See 501

UNIVERSITY OF MICHIGAN

S[illegible] 1955

SOCIAL SCIENCE LIBRARY

BF
1
J86

Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *AF Personnel and Training Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

UNIVERSITY OF MICHIGAN

SEP 26 1955

SOCIAL SCIENCE LIBRARY

CONTENTS

Multidimensional Stimulus Differences and Accuracy of Discrimination: C. W. ERIKSEN AND H. W. HAKE 153

Effect of Number of Permissible Response Categories on Learning of a Constant Number of Visual Stimuli: H. W. HAKE AND C. W. ERIKSEN 161

Influence of Frame Size on Apparent Length of a Line: T. M. KÜNNAPAS 168

A Comparison of Psychophysical Methods in the Investigation of Foveal Simultaneous Brightness Contrast: A. L. DIAMOND, H. SCHEIBLE, E. SCHWARTZ, AND R. YOUNG 171

Psychomotor Performance as a Function of Amount of Training and Stress: A. CASTANEDA AND D. S. PALERMO 175

Amount of Pretraining as a Factor in Stimulus Predifferentiation and Performance Set: J. H. CANTOR 180

Transposition in the White Rat as a Function of Stimulus Comparison: R. THOMPSON 185

Movement Error, Pressure Variation, and the Range Effect: B. WEISS 191

Generalized Extinction and Secondary Reinforcement in Visual Discrimination Learning with Delayed Reward: G. R. GRICE AND H. M. GOLDMAN 197

The Prediction of Decisions Among Bets: W. EDWARDS 201

Variations in Knowledge of Component Performance and its Effects upon Part-Part and Part-Whole Relations: E. A. BILODEAU 215

American Psychological Association

Vol. 50 No. 3 September 1955

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $7.50, or $15.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1955 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, Box 1557, Lackland Air Force Base, San Antonio, Texas. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

To save printing costs supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 [4, Suppl.], 389–449). In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Copyright, 1955, by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 50, NO. 3 SEPTEMBER, 1955

MULTIDIMENSIONAL STIMULUS DIFFERENCES AND ACCURACY OF DISCRIMINATION [1]

CHARLES W. ERIKSEN AND HAROLD W. HAKE

The Johns Hopkins University

Can stimuli be more accurately discriminated from one another if they differ on two or three dimensions than if they differ along a single dimension? And, can the discrimination accuracy for such multidimensional stimuli be predicted from the discriminability of the component dimensions? Little experimental work has been performed on these specific questions, but there are several reports in the literature that indicate multidimensional cues facilitate learning and reduce reaction time. Eninger (4) found that rats learned a discrimination problem more rapidly when visual and auditory cues were compounded than when either auditory or visual cues were used alone. Similar findings were obtained by Miller using compounds of visual and auditory modalities in a classical conditioning study (10). In both these studies compounding was across different sense modalities. However, Weitz (12) has shown that reaction time is reduced when two dimensions within the same sense modality are compounded. He found that reaction time to control levers was more rapid if the knobs differed in both hue and form than if they differed only in hue or only in form.

Several other studies have indicated, however, that the over-all perceptual situation determines in some cases whether or not compounded or multidimensional cues contribute to discrimination. Eriksen (5) found that compounding cues on the dimensions of hue, form, size, and brightness did not increase the speed with which objects could be located in a visual display. He found that this failure to improve location time was attributable to a perceptual heterogeneity factor that resulted from introducing more dimensions of variation into the display. More recently Bricker (1) has shown that redundant position cues were not useful in learning to associate labels to a set of visual forms.

In general, then, the results of work in this area have been to suggest that the effects of compounding dimensional cues depend in some measure upon the response studied and the nature of the task. In none of the cases has there been a direct effort to

[1] This research was performed under Contract No. AF 33(038)–22642 between the Institute of Cooperative Research, The Johns Hopkins University, and the Psychology Branch of the Aero Medical Laboratory, Directorate of Research, Wright Air Development Center.

assess the effects of compounding dimensions in terms of accuracy of discrimination or the number of discriminable steps in a stimulus series that can be obtained. In the present study both discrimination accuracy and the number of reliably discriminable steps have been determined for a series of stimuli varying along the dimensions of size, hue, and brightness. The discrimination measures obtained for stimuli varying along these single dimensions were then compared with the discrimination measures obtained when a stimulus series varied on size and hue, size and brightness, hue and brightness, and size, hue, and brightness.

Method

Design.—Discrimination was determined by the method of absolute judgment, or as it is sometimes called, the method of stimulus rating. The method requires *Ss* to identify by a name, number, or value each member of a set of individually presented stimuli. In the present study, *Ss* used the numbers from 1–20 in order to identify which stimulus had occurred. The absolute method was chosen as a measure of discrimination since it is amenable to an analysis in terms of information theory. Garner and Hake (7) have shown that the information transmitted by *Ss* about a series of stimuli along a dimension is interpretable as the number of stimuli that can be selected from this dimension to yield perfect discrimination under absolute conditions.

Seven different series of stimuli were judged in the present study. In three of these series the stimuli composing the series differed from one another along a single dimension. These dimensions were, respectively, size, hue, and brightness. In the remaining four stimulus series multidimensional stimulus differences were employed. In three of these four remaining series each stimulus in the series could be uniquely specified on each of two dimensions, and in the fourth series each stimulus could be specified on all three dimensions. The brightness and the size-brightness series each contained 17 stimuli, and 20 stimuli were used in each of the remaining series.

Each of six *Ss* made 100 judgments of each stimulus in each of the seven stimulus series. The *Ss* completed all judgments of a series before beginning their judgments on the next series.

TABLE 1

Specification of Stimulus Categories Used for Size, Hue, Brightness, and Hue-Brightness

Response	Stimulus Dimensions			
	Size (in. sq.)	Hue	Brightness	Hue-Brightness
1	1/8	5 R 5/6	N 1/	5 R 4/4
2	2/8	10 R 5/6	N 1.5/	10 R 5/4
3	3/8	5 YR 5/6	N 2/	5 YR 6/4
4	4/8	10 YR 5/6	N 2.5/	10 YR 7/4
5	5/8	5 Y 5/6	N 3/	5 Y 8/4
6	6/8	10 Y 5/6	N 3.5/	10 Y 7/4
7	7/8	5 GY 5/6	N 4/	5 GY 6/4
8	8/8	10 GY 5/6	N 4.5/	10 GY 5/4
9	9/8	5 G 5/6	N 5/	5 G 4/4
10	10/8	10 G 5/6	N 5.5/	10 G 3/4
11	11/8	5 BG 5/6	N 6/	5 BG 2/4
12	12/8	10 BG 5/6	N 6.5/	10 BG 3/4
13	13/8	5 B 5/6	N 7/	5 B 4/4
14	14/8	10 B 5/6	N 7.5/	10 B 5/4
15	15/8	5 PB 5/6	N 8/	5 PB 6/4
16	16/8	10 PB 5/6	N 8.5/	10 PB 7/4
17	17/8	5 P 5/6	N 9/	5 P 8/4
18	18/8	10 P 5/6	....	10 P 7/4
19	19/8	5 RP 5/6	....	5 RP 6/4
20	20/8	10 RP 5/6	....	10 RP 5/4

Note.—The specifications for hue and brightness are given in the Munsell notation. For example, 10 R 5/6 indicates that the hue is red with a brightness value of 5 and a saturation of 6 on the Munsell scales.

To control for practice effects, *Ss* made their judgments of the seven stimulus series in a counterbalanced order.[2]

Materials and apparatus.—Table 1 defines the stimuli used in the size, hue, brightness, and the hue-brightness series and shows the response number that was assigned to each. The other three multidimensional series are not represented in this table since they were constructed by compounding stimuli of the single dimensional series. In constructing these multidimensional series the component single dimension stimuli were superimposed one on top of the other. Thus for the series size-hue, Stimulus 1 was a 1/8 in. 5 R 5/6 square, Stimulus 2 a 2/8 in. 10 R 5/6 square, Stimulus 3 a 3/8 in. 5 YR 5/6 square, etc.[3] In

[2] The design followed a 7×7 latin square which was not quite realized owing to the seventh *S*'s failure to complete the experiment. However, statistical examination of the data revealed no evidence of systematic practice effects from one series to the next.

[3] The multidimensional series of hue-brightness was an exception to this method of construction and as a result has been presented in Table 1. This exception was necessitated by the nonindependence of hue and brightness and the inability to obtain Munsell hue in half brightness

constructing the multidimensional series of size-hue-brightness, the hue-brightness stimuli were compounded with the size stimuli in the manner described above for size-hue.

As can be seen in Table 1 the response numbers were assigned to stimuli in such a manner that the serial order of the response numbers conformed to the position of the stimuli along the dimension(s).

The stimuli for the size series were cut from a dark gray paper and for the hue, brightness, and hue-brightness series 7/8-in. squares of Munsell papers were used. For all series the stimuli were mounted in the center of 3-in. white cardboard squares to facilitate handling.

Procedure.—The *S*s were run individually with an average of five 1-hr. sessions being required to judge the stimuli in any one series. Prior to beginning the experimental judgments on a given series, each *S* made eight practice judgments with correction of each stimulus in the series. Following the practice judgments 100 judgments of each stimulus were obtained during the experimental sessions. In all, each *S* made a total of 13,400 experimental judgments over the seven stimulus series during the experimental sessions.

As was mentioned above, all stimuli were mounted on white cardboard squares. Four of these stimulus cards were constructed for each stimulus. The stimulus cards were presented to *S* one at a time and his judgment recorded. The experimental sessions were broken up into blocks of judgments where a block was defined as four judgments on each stimulus in the particular series. Between blocks *S* was given a short rest while *E* thoroughly reshuffled the stimulus cards in order to insure a random order of appearance of the different stimuli.

Subjects.—The six *S*s were male undergraduate students at The Johns Hopkins University. None had previously served as *S* in a psychophysical experiment. All were screened by a shortened version (3) of the Ishihara Color Vision test to insure that they were color normal.

Results

Two different methods have been used to express the discrimination accuracy obtained with the three single and the four multidimensional stimulus series. The first of these was to determine the information transmitted in bits (I_t) between the stimuli and responses in each series. Garner and Hake (5) have shown that the antilogarithm to the base 2 of I_t is interpretable as the maximum number of stimulus values that could be selected from a series of stimuli if errorless discrimination were desired. The second method consisted of computing the average percentage of correct responses given to the stimuli within each series.

As a first step in evaluating the results an over-all test of the significance of differences in discrimination accuracy among the different series was made. The value of I_t for each *S* was computed for each of the seven series. These I_t values were then subjected to a two-way analysis of variance (stimulus series and *S*s). The resulting F of 20.48 ($P > .0001$) demonstrated that the variation in accuracy between series was much

steps. It will also be noted in the table that the hues in this series have a saturation of 4 in the Munsell system while in the hue and the size-hue series all hues have a saturation of 6. Since the hues are of a different saturation and the brightness differences between adjacent stimuli are a full brightness step, the results obtained with this multidimensional series are not directly comparable to those obtained with the other multidimensional series.

TABLE 2

Discrimination Accuracy for the Three Single and the Four Multidimensional Stimulus Series

Stimulus Series	I_t	No. Stimuli Absolutely Discriminable	Average Per Cent Correct Judgments	
			Observed	Predicted
Size	2.84	7.19	47.5	
Hue	3.08	8.45	53.4	
Brightness	2.34	5.06	41.3	
Size-Hue	3.55	11.90	76.2	72.6
Size-Brightness	2.98	7.89	59.8	59.5
Hue-Brightness	3.76	13.55	84.2	—
Size-Hue-Brightness	4.11	17.28	96.5	90.7

Note.—Based upon error term for the analysis of variance, differences of .38 and .52 in I_t are significant at the .05 and the .01 level, respectively.

greater than could be reasonably attributed to chance.

The discrimination accuracy obtained with each of the seven series is shown in Table 2. The first column of this table gives the average value of I_t for the six *S*s. In the second column the antilog of I_t to the base 2 is shown. Thus the values in this column represent the best estimate of the maximum number of stimulus categories or values that could be selected from the different series if errorless discrimination were to be obtained.

As can be seen from Table 2, there is a marked gain in discrimination when two or three dimensions are compounded. This is true irrespective of which measure of discrimination is considered. The use of the multidimensional series of size-hue results in significantly more accurate discrimination than is obtained by either the size or the hue dimension used alone. The size-brightness and the hue-brightness series also yield better discrimination than is obtained by any of the component dimensions used singly; and, in the case of the series size-hue-brightness where stimuli differed from adjacent stimuli in the series on each of these dimensions, discrimination was nearly perfect for all 20 stimuli.

This improvement in discrimination obtained by compounding dimensions indicates at least a moderate degree of independence in the accuracy with which two or more dimensions are simultaneously judged in the same stimulus. That is, the improvement could be expected only in the event that *S* could simultaneously judge the stimuli on more than one of the component dimensions and that his judgments about values along these dimensions were not completely determined by his judgments of values along one of them. In the extreme case, for example, where *S*'s perception of values along all other dimensions in a series of compound stimuli was completely determined by his perception of values along just one of the dimensions, then his ability to discriminate among the compound stimuli could be no better than his ability to discriminate among stimuli differing only in the single dimension. Since accuracy of judgment for compound stimuli was considerably better in this experiment than that for stimuli differing on a single dimension, a considerable degree of independence of judgment along the component dimensions in the compound stimuli must be assumed.

TABLE 3

DISTRIBUTION OF RESPONSES IN PER CENT TO THE 8/8-IN. SQUARE IN THE SIZE SERIES, THE 10 GY IN THE HUE SERIES, AND TO THE COMPOUND OF THESE TWO VALUES IN THE SIZE-HUE SERIES
(N = 600)

Response	Size	Hue	Size-Hue Observed	Size-Hue Predicted
5	0.3	0.0	0.0	0.0
6	0.7	0.0	0.0	0.0
7	16.8	0.8	1.7	2.1
8	46.2	49.6	81.5	72.8
9	24.0	19.8	12.3	18.3
10	8.8	17.5	4.3	5.6
11	3.0	9.5	0.2	1.2
12	0.2	2.7	0.0	0.0

The results obtained with the multidimensional series can be predicted rather well by the use of three assumptions concerning discrimination behavior under the present conditions. They are:

1. At a given moment in time, the response tendency (rating) evoked by a magnitude on one dimension is independent of the response tendency evoked by magnitudes on other dimensions.

2. The distribution of response frequencies to a stimulus magnitude is a measure of the relative strengths of response tendencies elicited by that magnitude.

3. In the case of competing response

tendencies, the stronger tendency will be evoked as a response.

In Table 3 these assumptions have been used to predict the distribution of response to Stimulus 8 (8/8-in. 10 GY square) in the multidimensional size-hue series. The first column of this table shows the response distribution obtained for this size (8/8-in. square) when it occurred as Stimulus 8 in the size series and the second column shows the distribution of responses to this hue (10 GY) when it occurred as Stimulus 8 in the hue series. The third column is the response distribution to the stimulus compound of these two stimulus magnitudes or values.

If we consider the first two columns as reflecting the strengths of response tendencies to the size and hue magnitudes or values when viewed separately and assume also that the response tendencies evoked by each dimension are completely independent of each other, then we can expect that on 16.8% of the appearances of compound Stimulus 8 response tendency "7" would be evoked by the size of the stimulus and on 49.6% of its appearances response tendency "8" evoked by its hue. Since it is assumed that these two events are independent, we may expect that on about one-half (49.6) of the occasions on which *S* perceived the size of Stimulus 8 as being "7" a stronger tendency to respond "8" will be evoked by the hue of the stimulus. On all occasions in which these two events occur together the stronger tendency will appear as an overt response, and, by this means, about one-half of *S*'s underestimations of the size of the compound stimulus by one scale value will be corrected by his more accurate perception of the hue.

In a similar way we may consider all the other possible events that could arise from these two distributions. This essentially reduces to multiplying each response percentage for the size dimension by each of the response percentages for the hue dimension. The product in each case is assigned to the response category that has the highest percentage value. For example, .3% for Size Response 5 would be multiplied by .8% for Response 7 for hue and the product .24% would be assigned to Response 7 since it had the stronger response tendency.

Column 4 in Table 3 is the predicted response distribution for Stimulus 8 in the size-hue series. It was obtained by following the procedure which has been described through the 48 multiplications involved and then summing the products assigned to each response category. The predicted distribution is a fair approximation to the empirical one.

The distribution predicted in Table 3 is for only one of the compounded stimuli, however, and is not an adequate test of the predictive model. Accordingly, the procedure was carried out for each stimulus in the size-hue, size-brightness, and the size-hue-brightness series.[4] The predicted percentage of correct judgments for each stimulus was then averaged for each series. These average predicted percentages of correct responses are shown in Column 4 of Table 2 where they can readily be compared with the obtained values.

Comparison shows the predicted percentage of correct responses to be very close to the empirical results. The largest discrepancy is 5.8% for the multidimensional series size-hue-brightness, but it is less than 1% for size-brightness. The closeness of the fit can be further demonstrated by computing the rank-order correlation between the predicted percentage of correct judgments for each stimulus in a series and the obtained percentage of correct judgments. For the size-hue series this rank-order correlation is .86 for the 20 stimuli, and for

[4] It was not possible to obtain such predictions for the hue-brightness series since, as was pointed out in the section on procedure, this multidimensional series was not a simple compound of the stimuli used in the hue and brightness series. Owing to limitations of Munsell hues it was necessary to use a full brightness step between adjacent hues in this multidimensional series, whereas in the brightness series alone, half brightness steps were employed. Consequently, it was necessary for the same brightness values to be compounded with several different hues.

the 17 stimuli in the size-brightness series, rho is .77. A similar rank-order correlation for the size-hue-brightness was not practicable since the accuracy of discrimination for this series was nearly perfect and there was little or no variation among the predicted or obtained judgments of the stimuli.

Discussion

The results of the present study show quite clearly that compounding dimensional differences among stimuli results in a large and significant improvement in discrimination using the absolute method of judgment. In effect, the results show also that the small values for I_t, the maximum number of stimuli which can be discriminated with essentially no error, which have been obtained by using the absolute method (**6, 8**), are not necessarily the result of limitations imposed by the method itself but rather by the paucity of perceptual cues provided by stimuli varying on a single dimension. The very large numbers of different objects and events which are separately recognizable and identifiable in the world, then, may be at least partially explicable in terms of the multiplicity of cues available for discrimination.

Generalization from our results should be attempted with some caution, however, because not all perceptual tasks, even of the absolute judgment type, are identical with the tasks performed by our *S*s. The importance of the particular perceptual task and the performance measure studied has been pointed out earlier in this paper to be a factor in determining the effect of compounding dimensional differences between stimuli (**1, 5**).

It is important to recognize also that the improvement in judgmental performance found, when stimuli were made more complex by compounding two or more dimensions in the same stimulus series, resulted primarily from the ability of *S*s to make fairly independent judgments of stimulus values along each of the component dimensions. We cannot assume that this is always the case for compound stimuli. For some stimuli, judgments made of values in the separate dimensions may be interrelated. That is, the evocation of a particular response tendency by a component of a compound stimulus may change the likelihood of evocation of other response tendencies by other components of the stimulus. We could expect this to occur when *S*s have learned by long experience to associate the occurrence of certain values in one dimension with the occurrence of particular values in another. For example, we have assumed on the basis of the data in Column 1 of Table 3 that Stimulus 8 in the size-hue series would evoke response tendency "7" 16.8% of the time. If our *S*s had had any previous firm knowledge that lower-numbered sizes tended to occur with lower-numbered hues, we should not expect them to be as likely to assign higher numbers to the hue of the compound stimulus, given that they perceived the size to be "7," as they would if judging the hue in isolation. And, in that case we could not expect *S*s to correct their underestimations of size as often by making correct estimates of the hue values.

It is probable that most of the stimuli which we perceive outside the psychophysical laboratory involve such interdependencies among the several stimulus dimensions of which they are composed. A relationship between the area of a stimulus and its apparent intensity has been reported by Hanes (**9**). Other examples include the relationship between pitch and loudness (**11**) and the relationship between the proximity of adjacent parallel lines in the visual world and the likelihood that the lines are a part of a cohesive object in the world (**2**). We have included among our examples interdependencies which may be due to the nature of our receptor mechanisms and to the nature of our world. Both factors undoubtedly are concerned here.

Another very real consideration qualifying our results is the possible effect of

compounding a very large number of stimuli in a single stimulus series. This may result in a loss of ready awareness by *S*s of some of the dimensions involved. If this were true, we should expect there to be some upper limit to the amount of improvement in discrimination which can be obtained by compounding dimensions.

There is one other aspect of the present data that deserves comment. This concerns the number of absolutely identifiable or discriminable steps that can be obtained with the hue and brightness dimensions. Since the entire range of visible hue and nearly the entire range of brightness was used in these respective stimulus series, the number of absolutely discriminable steps obtained for these dimensions may be considered as close approximations to the maximum obtainable. As was seen above, very few values could be selected from the brightness dimension so as to satisfy the criterion of perfect discriminability under absolute conditions. The data suggest a maximum of about 5 values. The maximum number of values that could be selected from the hue dimension to meet this criterion is somewhat larger. Here the data indicate that at least 8 hues of equal brightness and saturation could be so selected. This number for hue is somewhat less than the number that has been estimated by Halsey and Chapanis (**8**). Their data indicated that between 10 and 12 hues could be absolutely discriminated.

This discrepancy between the two sets of data seems most likely to be due to differences in amount of practice given *S*s. Halsey and Chapanis practiced their *S*s until there was no further learning from trial to trial. If only the last 20 judgments of each of the 20 hue stimuli are considered in the present experiment, it is found that the *S*s were transmitting an average of 3.44 bits of information from the hue dimension. The antilog to the base 2 of this value is 11, which is quite comparable to the figure arrived at by Halsey and Chapanis. A comparable computation for the brightness series results in an average of 2.59 bits which indicates about 6 steps on the brightness dimension are absolutely discriminable.

Summary

The present study was concerned with the contribution of multidimensional stimulus differences to accuracy of discrimination. Discrimination accuracy was determined by the method of absolute judgment for a series of stimuli varying along the dimension of size, of hue, and of brightness. The discrimination measures obtained for these stimuli varying along these single dimensions were then compared with the discrimination measures obtained when a series of stimuli varied on size and hue, size and brightness, hue and brightness, and size, hue, and brightness.

The results demonstrated that the discriminability for a multidimensional series of stimuli was considerably greater than that obtained for any of the compounding dimensions used alone. It was further shown that the discrimination accuracy for a compounded or multidimensional series of stimuli could be predicted with reasonable precision if the discrimination accuracy of the compounding dimensions was known. This predictability was based upon the assumption of independence in the accuracy with which magnitudes on two or more dimensions are simultaneously judged.

References

1. Bricker, P. D. The effect of stimulus redundancy and response uncertainty on discrimination learning. *J. exp. Psychol.*, 1955, **49**, 73–82.
2. Brunswik, E., & Kamiya, J. Ecological cue-validity of "proximity" and of other gestalt factors. *Amer. J. Psychol.*, 1953, **66**, 20–32.
3. Chapanis, A. A comparative study of five tests of color vision. *J. opt. Soc. Amer.*, 1948, **38**, 626–649.
4. Eninger, M. U. Habit summation in a selective learning problem. *J. comp. physiol. Psychol.*, 1952, **45**, 604–608.
5. Eriksen, C. W. Object location in a complex perceptual field. *J. exp. Psychol.*, 1953, **45**, 126–132.
6. Eriksen, C. W., & Hake, H. W. Absolute judgments as a function of the stimulus range and the number of stimulus and response categories. *J. exp. Psychol.*, 1955, **49**, 323–332.

7. Garner, W. R., & Hake, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446–459.
8. Halsey, R. M., & Chapanis, A. On the number of absolutely identifiable spectral hues. *J. opt. Soc. Amer.*, 1951, **41**, 1057–1058.
9. Hanes, R. M. Suprathreshold area brightness relationships. *J. opt. Soc. Amer.*, 1951, **41**, 28–31.
10. Miller, J. The rate of conditioning human subjects to a single and multiple conditioned stimuli. *J. gen. Psychol.*, 1938, **20**, 399–408.
11. Stevens, S. S., & Davis, H. *Hearing.* New York: Wiley, 1938.
12. Weitz, J. The coding of airplane control knobs. In P. M. Fitts (Ed.), *Psychological research on equipment design.* Washington: U. S. Govt. Print. Off., 1947. Pp. 187–198.

(Received October 25, 1954)

Journal of Experimental Psychology
Vol. 50, No. 3, 1955

EFFECT OF NUMBER OF PERMISSIBLE RESPONSE CATEGORIES ON LEARNING OF A CONSTANT NUMBER OF VISUAL STIMULI [1]

HAROLD W. HAKE AND CHARLES W. ERIKSEN

The Johns Hopkins University

When *S*s are required to discriminate among a small set of previously unfamiliar visual stimuli, their ability to recognize each stimulus as a distinct member of the set can be increased by practice in verbally labeling the stimuli. The effects of such verbal training, during which *S*s learn to identify each stimulus with a unique irrelevant label, have been found in the enhancement of subsequent attempts of the same *S*s to learn a new set of responses to the same stimuli (**4**, **5**, **9**, **10**) or to discriminate among these same stimuli by a comparison method (**1**).

It is reasonable to expect a positive transfer effect to follow labeling practice because verbal training constitutes practice in discriminating among the stimuli. We should expect a gain whether the discriminating responses learned were verbal or some other set of motor responses. But it does not appear that the transfer effect results from the fact that *S*s are practicing the task of discriminating, as such. For example, we have no clear indication that transfer results because of factors suggested by Arnoult, such as "warm-up" effects or transfer of elements specific to the nature of the task or procedures rather than to the stimuli (**1**). We do know, however, that the transfer effect following verbal practice exceeds that which follows simply viewing the stimuli an equal number of times or even viewing the stimuli with the understanding that discriminative responses to the same stimuli would be required at a later time (**5**).

The fact that the perceptual gain following formal labeling practice exceeds that which follows simply viewing the stimuli is explicable from two points of view. The first stresses the importance of the information which is transferred to *S* on each trial of labeling practice. After *S* identifies the stimulus presented on each trial, he is routinely informed about the correctness of his identification or is told the arbitrarily predetermined correct label for that stimulus. In this way *S* can avoid confusing the stimuli perceptually because he has the opportunity to compare his own subjective identifications of the stimuli with a set of perfectly accurate external criteria. One advantage of learning verbal labels for anything may be that it allows us to compare the distinctions which we make among the objects of the world with those made by others.

If the information supplied to *S* on each trial is important in this way, then the transfer effect resulting from practice should depend upon the specificity with which he can identify the stimuli during practice. Where the number of labels used is a fraction of the number of stimuli in the set, for example, then *S* cannot state on

[1] This research was done under Contract No. AF 33 (038)–22642, between the Institute for Cooperative Research, The Johns Hopkins University, and the Psychology Branch of the Aero Medical Laboratory, Directorate of Research, Wright Air Development Center.

each trial which stimulus has appeared. He can only indicate that he perceives the stimulus as one of a group sharing the same label. This is true also of the information supplied on each trial, following the labeling attempt, which could indicate only the correct general label for the stimulus which had appeared. When the specificity of labeling is decreased, therefore, the transfer effect should decrease also because the external criteria with which *S* can compare his own identifications of the stimuli during practice are more ambiguous.

The same conclusion can be drawn from the second point of view, the notion that the critical aspect of verbal labeling practice is the acquisition and retention of the arbitrary relations between labels and stimuli (**4**). When *S* is allowed to use labels during verbal practice which are fewer in number than the stimuli, we should expect the specificity of the label-stimulus associations formed during practice to decrease and the reinforcement of attention to cues unique to each stimulus to be more ambiguous. And, if distinctive labeling has cue-producing properties as suggested by Miller (**8**), then, where a single label corresponded to a subgroup of stimuli, fewer cues would be available for discriminating among the stimuli at the end of practice. Experimental attempts to support this latter notion have produced conflicting evidence (**3, 9**).

Whatever the mechanism involved, it appears clear that if the degree of stimulus differentiation achieved during practice is closely related to the ability of *S* to state explicitly which stimulus appears on each practice trial, then the positive transfer effect resulting from practice should be decreased by any practice method which restricts this ability. In the experiment reported here we restricted the ability of *S* to specifically identify the stimuli during a practice session. This was done by forcing him to use a set of responses which was smaller in number than the set of stimuli to be labeled. He was required to learn to apply either 2, 4, or 8 labels to a set of 16 previously unknown patterns. An orthogonal design permitted an evaluation of the degree to which the resulting transfer effects were specific to the labeling tasks experienced in original practice and in a subsequent testing situation. In addition, the relationship between learning performances in the two phases of the experiment was studied in terms of individual *Ss* within groups and for individual stimuli as labeled by individual *Ss*.

Method

Apparatus. The visual patterns which *Ss* identified consisted of patterns of lights appearing on a 5 × 5 matrix of 25 small neon bulbs. The matrix was 7 in. on a side, and *Ss* faced it at a distance of 8 ft. Each of the 16 patterns consisted of five lighted bulbs which were selected at random from the matrix and which had no apparent meaning or structure when viewed together. The appearance of the patterns was controlled by an electronic sequence programmer which read the patterns from Teletype tape and presented them automatically on the matrix. The successive appearances of the patterns were spaced 10 sec. apart and each appearance lasted 2 sec.

Procedure.—All *Ss* saw each of the 16 patterns 20 times in a total of 320 trials. The patterns appeared in a partially restricted random order which provided for the appearance of all 16 patterns in the first 16 trials, in the second block of 16 trials, etc. During each trial and immediately following the appearance of the pattern, *Ss* were required to call out a label for that pattern. Each *S* had a card in front of him on which was printed a list of the labels he was permitted to use in labeling the patterns. The labels consisted of letters of the alphabet chosen to have as little sequential connection as possible. Immediately after *Ss* responded with a label on each trial, they were told the arbitrarily predetermined correct label for the pattern which had appeared on that trial. The task of each *S* was to learn which of the labels from his list could be correctly applied to each of the patterns.

TABLE 1

MEANS AND *SD*'S OF FREQUENCY OF CORRECT IDENTIFICATIONS PER BLOCK OF 16 TRIALS

No. of Labels in Learning Task	Learning Task							Transfer Task						
	Number of Labels to be Used in the Transfer Task							Number of Labels Used in the Transfer Task						
	2		4		8		Task Mean	2		4		8		Task Mean
	Mean	*SD*	Mean	*SD*	Mean	*SD*		Mean	*SD*	Mean	*SD*	Mean	*SD*	
2	11.06	3.20	10.80	3.90	11.16	4.72	11.01	11.54	3.44	9.44	6.74	8.99	4.93	9.99
4	6.36	3.36	6.80	4.28	7.08	3.25	6.75	11.62	4.87	8.89	4.13	7.06	6.43	9.19
6	5.89	5.90	4.84	5.21	5.88	6.07	5.54	12.51	5.90	8.18	5.43	8.13	8.82	9.61
Task mean	7.77		7.48		8.04			11.89		8.84		8.06		

The experimental procedure consisted of two tasks for all *Ss*: an original learning task of 160 trials during which they identified the 16 patterns using one set of labels, and a transfer task during which they identified the same patterns using another set of labels. Nine groups of *Ss* were used, each of which used a different set of labels in the two tasks. Assignment of sets of labels to the nine groups was determined by a factorial design in which the 10 *Ss* in each of the nine groups was assigned one of the nine combinations of the three sets of labels which could be used in the learning task and the three sets of labels which could be used in the transfer task. This design provided 30 *Ss* who used just 2 labels in the learning task, 30 *Ss* who used 4 labels in the learning task, and 30 *Ss* who used 8 labels in the learning task. Of the 30 *Ss* using a particular number of labels in the learning task, 10 later used 2 labels, 10 later used 4 labels, and 10 later used 8 labels in the transfer task. In all cases, including those *Ss* who used the same number of labels in the two phases, the labels used in learning were not those used in the transfer task. There was no meaningful relationship between the sets of labels used in the two tasks. Wherever possible the stimulus patterns which had shared the same label during the first task were separated and assigned to different labels during the second task.

Subjects.—A total of 90 *Ss* was used, 10 in each of the nine experimental groups. They were men undergraduate students who were run through the procedures one at a time.

RESULTS

The data of this study are the average numbers of correct identifications per block of 16 trials made by *Ss* during all trials of the learning and transfer tasks. A correct identification was recorded whenever *S* responded to a visual pattern with the arbitrarily predetermined correct label for that pattern.

The data were analyzed first to determine the effect upon performance of the number of labels used by *Ss* in identifying the patterns during original learning. The study of the effect of this variable upon performance during the transfer task was one of the main objectives of the experiment.

The effect of response restriction upon performance.—The average numbers of correct identifications per block of 16 trials averaged over all learning trials are presented in the left half of Table 1. The same measures of performance for the transfer task are shown in the right half of this table. The groups were assigned to the rows according to the number of labels used in original learning, and average performance in the rows provides an estimate of the effect of this variable upon performance during each task. The performance of each group of *Ss* in the first task is roughly proportional to the probability of chance success associated with the number of labels used in that task, although the

performances of *Ss* using 4 and 8 labels are more similar than such a simple notion would predict.

Average transfer performance in each of the rows of the right half of Table 1 measures the effect of the number of labels used in original learning upon performance during the transfer task independent of the nature of the transfer task. This effect appears to be slight. The *Ss* who used 2 or 8 labels in the learning task did slightly better during the transfer task than *Ss* who had used 4 labels.

The data of Table 1 were analyzed separately for each task by an analysis of variance technique after the group variances of each task had been tested for heterogeneity by Bartlett's test (**2**) and could be assumed to be homogeneous. Variation among the row means of Table 1 for the learning task was found to be significantly large ($P < .001$), as expected, because of the differences among the possibilities of chance successes associated with each number of labels. For the same reason there was a significant variation among the column means of the transfer data of Table 1. There were no other significant effects. The effect of the number of labels used in original learning upon performance during the transfer task was not reliable, nor was there any indication that the numbers of labels used in original learning interacted with the numbers of labels used in the transfer task.[2]

While analysis of the data provided no evidence that the nature of experience with the patterns during learning differentially affected transfer performance, the evaluation of the groups according to the number of labels used originally ignores another original learning variable which was not constant from group to group. This is the number of times *Ss* were correct in using the permissible labels in the first task. The different probabilities of chance success associated with the various numbers of labels used produced a significant variation from group to group in the number of correct labeling responses achieved in the initial labeling task. In effect, we have an unavoidable inverse relationship in both tasks between the number of labels permitted and the number of occasions in which the labels were applied correctly to the patterns. This relationship prevented a separate analysis of the effect of these two factors upon transfer performance.

The correctness or incorrectness of labeling responses during the learning task should be a critical aspect of *S*'s experience with the patterns if differentiation of the patterns during that task resulted from the act of learning a label for each pattern. Presumably, the correctness of labeling responses indicates the relative strengths of the specific stimulus-response associations built up during practice. We should expect to find considerable consistency of labeling performance in the two tasks if the associations formed during learning were directly related to ability to differentiate the patterns.

The consistency of labeling performance in the two tasks of the experiment was studied *within* each of the nine groups of our experiment. Two types of relationships were studied. The first was a measure of the consistency of performance of *Ss* in each group. This was the correlation between average performance in the learning and transfer tasks of the 10 *Ss* in each group. The second type of relationship measured the consistency of performance within-subjects. This was the correlation within each *S* between the number of correct responses made to each pattern in the learning and the transfer tasks.

We expected that these relationships would not be the same within all groups. This expectation was based upon the reasonable assumption that the degree to which *Ss* using 2 labels needed to distinguish among the patterns in order to label them correctly during learning was less than that required of *Ss* using 4 or 8 labels. With only 2 labels the effects of chance could contribute as heavily in producing correct responses to the patterns as could real ability to discriminate among them. The number of correct identifications achieved during learning with 2 labels could be expected, then, to be a less important indicator of the extent to which *Ss* could discriminate among the patterns than it

[2] The same analysis was performed after correcting for chance success. This removed all sources of significant variation. The effect of the number of labels used in learning upon the scores of the transfer task, after correction for chance, was still an insignificant source of variation.

would be for *Ss* using larger numbers of more specific labels.

Relationship between learning and transfer performance.—Coefficients of correlation were computed from the average learning and transfer performances of the 10 *Ss* in each of the nine groups. These are the within-group coefficients presented in Table 2. Of special interest are the average within-group correlations computed from the pooled variates of the three groups of *Ss* who used a particular number of labels originally, where each deviation involved was measured from its own group mean (**11**). These coefficients based upon the pooled data indicate some tendency for the correlation between learning and transfer performance to be reduced for *Ss* using the less specific labels originally. These differences are not statistically reliable, however, and the within-group regression can be assumed the same for all groups. The average within-group correlation of .57 for all groups properly can be taken as the best estimate of the relationship within all groups (**11**). Accounting for the average linear within-group regression results in a 33% reduction in the within-group variance, which is a statistically significant reduction ($P < .01$).

TABLE 2

CORRELATION AND REGRESSION COEFFICENTS FOR THE NINE GROUPS

Within-group coefficients computed from the average learning and transfer scores of 10 *Ss* in each group; within-*S* coefficients computed from average learning and transfer scores for individual patterns within *Ss*.

Groups	Within-Groups		Within-Ss*	
	Correlation	Regression	Correlation	Regression
2–2	.58	.62	.24	.19
2–4	.56	.97	.16	.16
2–8	.44	.46	.20	.27
Pooled data**	*.50*	*.66*	*.19*	*.20*
4–2	.25	.36	.33	.25
4–4	.39	.38	.12	.11
4–8	.30	.59	.13	.15
Pooled data**	*.30*	*.43*	*.19*	*.18*
8–2	.63	.66	.28	.19
8–4	.64	.67	.19	.18
8–8	.90	1.30	.18	.18
Pooled data**	*.74*	*.90*	*.21*	*.18*

* Within-*S* coefficients are based upon pooled variates from the 10 *Ss* in each group, each deviation involved being measured from its own *S* mean (**11**).

** These coefficients are based upon pooled variates from the 30 *Ss* using the same number of labels in the learning session, each deviation being measured from its own group mean or own *S* mean (**11**).

The significant relationship within the groups demonstrates a consistency of performance in the two tasks. The *Ss* who did well in learning did well in the transfer task also. But the implication of this reliable relationship is not straightforward because it could reflect merely the individual differences among *Ss* in ability to perform the labeling tasks. To the extent that the two labeling tasks demanded common skills we could expect those *Ss* who did well in learning to do well on the transfer task also. This may account for between-task correlations reported before in the context of verbal pretraining (**1**).

More specific information about the relationship between transfer performance and the number of correct identifications achieved during learning is provided by the within-*S* relationships. An evaluation of the relationship between performance in the two tasks within each *S* indicated the extent to which the patterns that *S* labeled successfully in transfer were those that he had labeled most successfully originally.

The within-*S* regression and correlation coefficients are presented in Table 2 as average within-*S* coeffi-

cients for each of the nine groups. These coefficients are based upon pooled values from the 10 *Ss* in each group, each deviation involved being measured from its own subject mean (11). Analysis of the small differences among the nine average within-*S*, within-group regression coefficients indicated that they were not significantly large. The average within-*S* correlation for all *Ss* is .20.[3]

Accounting for the average within-*S* regression does significantly reduce the transfer task within-subject variation ($P < .01$). The predictive value of this relationship is quite low, however, as the resulting reduction of within-*S* variation is only 4%. This is decidedly a very small effect compared to that we should expect to find if the number of correct identifications were the critical variable determining differentiation of the patterns prior to the transfer phase. Most of the variability which each *S* showed in labeling the 16 patterns during the transfer task must be attributed to factors other than the relative success with which *S* labeled each pattern originally.

Discussion

The results are not in accord with our notion that the number of labels used during initial verbal practice should determine the specificity or accuracy with which *Ss* learn to recognize each pattern during such practice. The use of two labels by one-third of our *Ss* in original learning allowed them only to assign explicitly the patterns to one of the two subsets corresponding to the two labels. Their transfer performance, however, was at least as good as that of *Ss* who could explicitly identify the patterns as belonging to one of eight subsets during the original task.

Contrary to our expectation also, the success with which *Ss* used the labels in the first task, regardless of the nature of the labels used, does not appear related in a critical way to the perceptual gains which resulted from labeling practice. The performance of *Ss* in the transfer task was related to performance during original learning to a significant degree, but we have reason to assume that transfer performance would be related to the same or greater extent to other tests of paired-associate memory or perceptual speed which did not use our stimulus patterns. Attneave has demonstrated a negative correlation of .64 between scores on two tests of the Airman Classification Battery (6 pp. 248–251, 376–382) and errors in labeling visual forms.[4] If we choose, we may regard performance in the learning task only as a measure of some pertinent abilities which each *S* brought to the experiment rather than as a measure of the degree to which he learned the stimuli during that session.

The *Ss* undoubtedly learned to differentiate the stimuli during the first task, but we cannot unequivocally attribute what they learned to the act of learning the correct label for each stimulus. The number of times each stimulus pattern was correctly paired with its appropriate label in the first task, regardless of the number of labels used, was not importantly related to the success with which that pattern was discriminated from all others in the set and correctly labeled in the transfer task. This appears to minimize the role that the specific stimulus-response associations formed in the first task could play in determining what is learned about the stimuli in that task.

The results appear to us to emphasize the importance of the general labeling *process* rather than factors related to the

[3] Both the within-group and the within-*S* relationships can be assumed linear in form. The addition of quadratic terms to the average regression equations did not result in significant reductions in the variances of the transfer scores under those achieved by fitting linear regression lines to the data (11).

[4] Reported in an unpublished manuscript entitled "Symmetry, information and memory for patterns."

particular labels used. We may judge from our results and from others (5) that the perceptual gain resulting from labeling practice appears to occur so long as *Ss* have a decision to make about the stimuli on each trial. The labeling task given to our *Ss* seems merely to have provided a context which defined an objective for them. Whatever the specific nature of this objective, it required that *Ss* learn to discriminate among the stimuli, and this appears to have been accomplished to some degree by all *Ss*, independently of the nature of the labeling task assigned.

Summary

The *Ss* were required to learn to apply either 2, 4, or 8 irrelevant verbal labels to a set of 16 similar and previously unfamiliar complex visual patterns. Immediately following practice in this task, *Ss* were required to learn to identify the patterns, again using a different set of irrelevant verbal labels that were either 2, 4, or 8 in number.

The performance of *Ss* on this transfer task was significantly affected by the number of labels used on this task but not by the number of labels used in the initial learning task. The extent to which each *S* correctly labeled each pattern during learning was not indicative to any important degree of his success in labeling each pattern during the transfer task.

The performance of each *S* during learning, relative to the other *Ss* in each group, was indicative of relative performance within each group on the transfer task. It was suggested that this relationship may be indicative merely of the extent to which both tasks demanded common skills. The extent to which *Ss* learned to differentiate the stimuli during learning was not indicated unequivocally by measures of performance in that task, and the perceptual gain which occurred from labeling practice was independent of the nature of the labeling task assigned.

References

1. Arnoult, M. D. Transfer of predifferentiation training in simple and multiple shape discrimination. *J. exp. Psychol.*, 1953, **45**, 401–409.
2. Bartlett, M. S. Some examples of statistical methods of research in agriculture and applied biology. *Suppl. J. Roy. Statist. Soc.*, 1937, **4**, 137–183.
3. Birge, J. S. The role of verbal response in transfer. Unpublished doctor's dissertation, Yale Univ., 1941.
4. Gagné, R. M., & Baker, K. E. Stimulus pre-differentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439–451.
5. Goss, A. E. Transfer as a function of type and amount of preliminary experience with task stimuli. *J. exp. Psychol.*, 1953, **46**, 419–428.
6. Guilford, J. P. (Ed.) *Printed classification tests.* Washington, D. C.: U. S. Govt. Print. Off., 1947. (AAF Aviat. Psychol. Program Res. Rep. No. 5, pp. 248–251.)
7. McAllister, D. E. The effects of various kinds of relevant verbal pretraining on subsequent motor performance. *J. exp. Psychol.*, 1953, **46**, 329–336.
8. Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.
9. Robinson, J. S. The effect of learning verbal labels for stimuli on their later discrimination. *J. exp. Psychol.*, 1955, **49**, 112–115.
10. Rossman, I. L., & Goss, A. E. The acquired distinctiveness of cues: the role of discriminative verbal responses facilitating the acquisition of discriminative motor responses. *J. exp. Psychol.*, 1951, **42**, 173–182.
11. Snedecor, G. W. *Statistical methods.* (3rd Ed.) Ames, Iowa: Collegiate Press, 1940.

(Received October 25, 1954)

Journal of Experimental Psychology
Vol. 50, No. 3, 1955

INFLUENCE OF FRAME SIZE ON APPARENT LENGTH OF A LINE

THEODOR M. KÜNNAPAS

University of Stockholm

In his classical investigation Rubin (**3, 4**) found that a visual perception is normally divided into two fields: a figure, which is usually the object of attention, and a ground, which occupies the rest of the perception. This distinction becomes especially clear in ambiguous fields as, e.g., in the vase-profile figure. He pointed out the phenomenal difference between the figure and the ground. The figure looks like a thing (*Dingcharakter*), whereas the ground appears as a substance (*Stoffcharakter*). The figure seems to be nearer to *O* than the ground. The contour belongs to the figure and not to the continuous ground.

The figure-ground problem is developed further by gestalt psychologists, in the main by Kurt Koffka (**2**). He divided the behavioral environment into things and frameworks. Koffka pointed out that the figure-ground articulation is a part of that more general problem, things and frameworks. Figures are things, and grounds operate like frameworks in perception. He says (**2**, p. 184): "The figure depends for its characteristics upon the ground on which it appears. The ground serves as a framework in which the figure is suspended and thereby determines the figures." He demonstrated excellently the framework character of the ground by its influence on the shape of the figure.

In the present investigation an attempt is made to study how the frame size influences the apparent length of a line inclosed within a frame. The experimental situation is simplified: the line is one of the simplest figures as well as the square is one of the simplest frames. In this way some more exact quantitative formulations about the influence under study may be expected.

METHOD

Stimulus materials.—White square pieces of cardboard were used as frames. The lengths of the sides of these squares were 7, 9, 12, 16, and 21 cm. The length of the constant square was 7 cm., and the other lengths refer to the variables that were compared with it. In the middle of each square a line was drawn. The length of the line in the constant square was always 50 cm., while in the variable squares it varied. The interval of variation was 1 mm. The width of the lines was .4 mm.

Procedure.—White square pieces were placed on the very large sky-blue background, only two at a time—the constant square and a variable. The variable square was placed either to the right (R) or to the left (L) of the constant. The variable square was always placed on the background so that the line at the middle was at the same height as the line in the constant. The distance between the two lines was always 13 cm. Consequently the distance between the two squares on the background was variable: the largest square was placed 4 cm. from the constant square, the other squares were 6.5, 8.5, and 10 cm. (the smallest square) distant from it.

The two lines that were compared were situated 30° below eye level. The angle of exposure of the squares was 60°. The distance between *S*'s eyes and the lines was 110 cm.

The constant method was employed. After preliminary trials, for every size of square a limited number of lines were selected which were presented to *S*s in a prearranged order that was unknown to them. The *S* had to compare the length of the exposed two lines in two squares, and report whether the variable line was shorter, equal to, or longer than the constant line. A maximum of 5 sec. was allowed for making the judgments. The cases where *S* could not make a definite report during this time were counted as "equal."

TABLE 1

LENGTH OF LINE (IN MM.) WHICH APPEARS EQUAL TO THE CONSTANT LINE OF 50 MM. IN THE CONSTANT SQUARE

Size of Square (Cm.)	PSE	L_1	L_2	IU
7×7	(50.00)			
9×9	51.33	52.53	49.69	2.84
12×12	52.68	54.19	51.01	3.18
16×16	54.16	55.37	52.43	2.94
21×21	55.63	57.14	53.98	3.16

Each size of the variable squares was used 20 times on two different days: 10 L, and 10 R positions. The sequence of the experiments was rotated so as to equate any possible learning and fatigue effects.

Subjects.—Ten students, five men and five women, took part in the experiments. All *S*s had normal visual acuity.

RESULTS

For each of the four different sizes of the square, the point of subjective equality (PSE) was taken as the significant measure of *S*s' responses. Also the upper limen (L_1) and the lower limen (L_2), as well as the interval of the uncertainty (IU), were computed by Spearman's method of the arithmetic mean (1).

Table 1 gives the general results. There is no doubt that the apparent length of the line which is in the middle of a square-shaped frame depends on the size of this square. The larger the square, the longer the line must be physically if its length is to appear equal to that of the constant line of 50 mm. in the middle of the constant square. Consequently an increase in the size of the frame reduces the phenomenal length of the line drawn on it, just as if it were compressed.

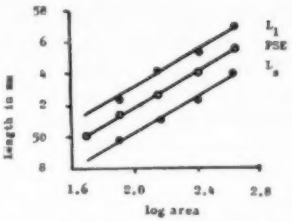

FIG. 2. Length of the line which appears equal to the constant line of 50 mm., plotted against the logarithm of the area of the square-shaped frame.

Figure 1 shows that the physical length of the line which appears equal to the constant line of 50 mm. does not increase rectilinearly. The curve is rectilinear, however, if plotted against the logarithm of the area of the square-shaped frame (see Fig. 2). Thus, the phenomenal length of the line is a logarithmic function of the area of the square enclosing it.

Figure 3 indicates that the length of the line likewise increases recti-

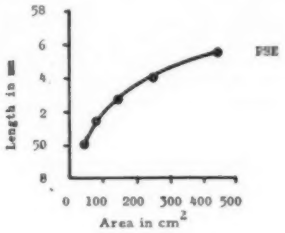

FIG. 1. Length of the line which appears equal to the constant line of 50 mm., plotted against the area of the square-shaped frame.

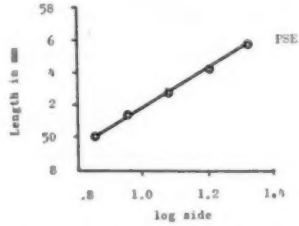

Fig. 3. Length of the line which appears equal to the constant line of 50 mm., plotted against the logarithm of the side of the square-shaped frame.

linearly, if plotted against the logarithm of the side of the square enclosing it. This means that the phenomenal length of the line is also a logarithmic function of the side of the square-shaped frame.

In reality these two functions are the same if we are dealing with a square. According to Fig. 2, $L = k \log A$, where L = phenomenal length of the line, k = constant, and A = area of the square. But in a square, $A = S^2$, where S = side of the square. Substituting S^2 for A, we obtain $L = 2k \log S$, or $L = c \log S$, where $c = 2k$.

Also *Ss*' individual responses show the same relation. In Fig. 4 the responses of three *Ss* are given: *a* represents one of the best fittings, *b* represents a typical curve, and *c* represents one of the atypical curves. In *c* the deviations from the rectilinear increase are greatest.

Observations made during the course of the experiments indicate that in the line-frame articulation the frame influences the line not only in a horizontal but also simultaneously in a vertical direction. The frame reduces both the length and the width of the line at the same time. The lines, which were of the same width (.4 mm.) in all the squares, seemed to have different width according to the different size of the square-shaped frames. In the larger squares the lines appeared to be thinner, while in the smaller squares they seemed to be thicker, i.e., the influence of the frame on the width of the line has the same direction as on the length. This difference was so noticeable that some *Ss* asked sometimes whether the lines really were of the same width. Thus we may say that the frame influences the enclosed line in all main directions at the same time. There is also reason to assume that the same influence occurs in the more complex figure-frame situations.

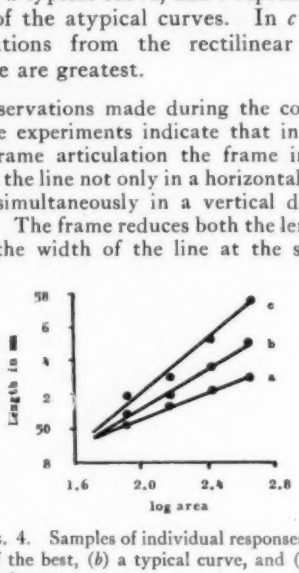

FIG. 4. Samples of individual responses: (*a*) one of the best, (*b*) a typical curve, and (*c*) an atypical curve.

SUMMARY

The present study is concerned with how frame size influences the apparent length of a line inclosed within the frame. The square, as one of the simplest geometrical figures, served as frame. Five different sizes of the square were employed. The lengths of the lines in the middle of these squares were compared with the constant line of 50 mm. on the constant square 7 × 7 cm.

The results indicate that the frame greatly influences the apparent length of the line. The phenomenal length of the line is a logarithmic function of the area and the side of the square-shaped frame. Observations made during the course of the experiments indicate that in the line-frame articulation the frame influences the line not only in a horizontal direction but also probably in a vertical direction.

REFERENCES

1. GUILFORD, J. P. *Psychometric methods.* (2nd Ed.) New York: McGraw-Hill, 1954.
2. KOFFKA, K. *Principles of gestalt psychology.* London: Routledge & Kegan Paul, 1935.
3. RUBIN, E. *Synsoplevede Figurer.* Köbenhavn: Gyldendalske Boghandel, 1915.
4. RUBIN, E. *Visuell wahrgenommene Figuren.* Köbenhavn: Gyldendalske Boghandel, 1921.

(Received October 14, 1954)

Journal of Experimental Psychology
Vol. 50, No. 3, 1955

A COMPARISON OF PSYCHOPHYSICAL METHODS IN THE INVESTIGATION OF FOVEAL SIMULTANEOUS BRIGHTNESS CONTRAST [1]

A. L. DIAMOND, H. SCHEIBLE,[2] E. SCHWARTZ[3], AND R. YOUNG

Northwestern University

Investigations of simultaneous brightness contrast have shown that adjacent visual fields can affect one another's apparent brightness. In such investigations the field in which the changes in apparent brightness are measured is called the test field. The adjacent field is termed the inducing field.

The technique usually employed to measure the effect of the inducing field upon the test-field apparent brightness involves a binocular match (**1, 2, 4, 5**). That is, *S* is asked to equate the apparent brightness of the test field, presented to one eye, to the apparent brightness of a standard or match field, presented to his other eye. In this way, when the test-field apparent brightness changes as a function of some variation in the inducing field (presented to the same eye as the test field) the brightness equality match, between match and test fields, indicates the extent of this change.

Further, as in the above-cited studies, the psychophysical method employed is usually that of adjustment, as described by Graham (**3**). In this method, *S* varies either the match-field or the test-field luminance in order to obtain a brightness equality match between the two. The fact that *S* controls the luminance changes necessary in the brightness match, therefore, enables him to utilize such motor cues as might arise from his manipulation of the apparatus, in addition to the experimentally intended visual cues of test- and match-field luminances.

It is conceivable then that certain aspects of the contrast phenomenon might be peculiar to the particular experimental method used,—the method of adjustment in this case. The purpose of the present paper is to study simultaneous brightness contrast using three different psychophysical methods to determine: (*a*) whether the characteristic effects seen in contrast are dependent upon a particular investigating method; and (*b*) which method is the most useful and reliable in studying contrast.

This is tested in three separate experiments, in each of which two psychophysical methods are compared as follows: Exp. I compares the method of limits to that of constant stimuli; Exp. II compares the method of adjustment to that of constant simuli; and Exp. III compares the method of limits to that of adjustment. Each of these methods is generally described by Graham (**3**).

METHOD

General Method

Apparatus.—The apparatus used in all three experiments is identical to that described by Diamond (**1**). It provides as a measure of test-field apparent brightness the luminance of a match field, which can be set, either by *E* or *S*, equal to the test-field apparent brightness. The test and inducing fields are adjacent squares, each 33′ of visual angle on a side, presented to

[1] This work was supported in part by a research grant NSF-G923 from the National Science Foundation and in part by the Graduate School of Northwestern University.

[2] Now at the University of California, Los Angeles.

[3] Now at DePaul University, Chicago.

S's right eye, which fixates a 3′ point of light located 21′ to the left of the test field. The match field, also a 33′ square, is presented to *S*'s left eye which fixates a 3′ point of light located 21′ to the right of the match field. The *S* is instructed to fuse left- and right-eye fixation points so that in the binocular view one fixation point is seen between the match field on the left and the test field, underneath the inducing field, on the right. The visual angle between the fused fixation point and the farthest corner of any field is 70′. Stimulation, therefore, is essentially foveal.

Procedure.—The common procedure was as follows: Seated in a light-tight cubicle, *S* initially dark adapted for 3 min., then light adapted for 3 min. to (*a*) the test field, always at 0.6 log ml.; (*b*) the match field; and (*c*) the inducing field, its luminance being zero at the outset of each session. Then, under each of the nine conditions of inducing-field luminance (varied in order from $-\infty$ to 2.71 log ml.) *S* was given four practice matches followed immediately by a set of experimental matches. There were 10 such experimental matches per condition under the methods of adjustment and limits, and 50 under constant stimuli. One minute of dark adaptation and 1 min. of light adaptation came between the successive inducing-field luminance conditions. In each of the three experiments a different E and different *S* were used.

Experiment I

The methods compared in this experiment were that of limits and constant stimuli. Two *S*s, AL and DR, were given both methods. AL received limits the first session and constant stimuli the second, DR took these in reverse order.

Method of limits.—To obtain an apparent brightness judgment, *E* first increased the match-field luminance in small steps, from that point at which *S* reported its apparent brightness to be definitely less than that of the test field until *S* judged the match field to appear just noticeably brighter than the test field. Five increasing and five decreasing judgments were made, in counterbalanced order (ABBA . . .) and the mean of ten match-field luminances corresponding to these judgments was taken as a measure of the test-field apparent brightness.

Method of constant stimuli.—The *E* first determined a range of match-field luminances the extremes of which appeared to *S* as just dimmer and just brighter than that of the test field. Then, in this range he selected five equally spaced luminances, setting the match field at each luminance value in a random manner until he had presented each value 10 times. At each individual setting, *S* reported whether the match field appeared brighter or dimmer than the test field. The *E* then blocked the match-field light source and changed its luminance to another value. Since this change required a maximum of 8 sec., the match field was exposed for 8 sec. during each individual judgment in order to keep the left eye at a constant level of adaptation. The percentage of occurrence of the response "brighter" was then graphed for the five match-field luminance values, and that luminance which coincided on the graph with the 50% point along the ordinate (5 out of 10 responses being "brighter") was taken as the average measure of the test-field apparent brightness.

Experiment II

The methods of adjustment and constant stimuli were compared in this experiment. Two *S*s, SM and BM, were given both methods in counterbalanced order.

Method of adjustment.—In making his brightness equality judgments, *S* (rather than *E*) set the luminance of the match field so that it appeared equal to that of the test field. The *E* then changed the match-field luminance in a random direction after each match. The *S* had to adjust the match-field luminance again until it seemed equal in apparent brightness to that of the test field. The mean of 10 such match-field luminance settings was taken as a measure of the test-field apparent brightness.

Method of constant stimuli.—This is identical to the same method described above.

Experiment III

In this experiment the methods of limits and adjustments were compared for three *S*s. The first two, PC and MG, first had adjustment, then limits. The third, PS, took the methods in reverse order. These methods are identical to the same methods described above.

Results

With regard to the general applicability of the methods tested, one result is clear. The method of constant stimuli is much more time consuming and laborious than the other two methods. Experimental sessions under constant stimuli took from 90 to 120 min. to collect a comparable amount of data collected in 45 to 60 min. under the other methods. In addition, under constant stimuli, *S*s complained of eye strain, headaches, and fatigue, and *E*s of fatigue, whereas

TABLE 1

TEST-FIELD APPARENT BRIGHTNESS (B_m) AS A FUNCTION OF INDUCING-FIELD LUMINANCE (B_i)

(Data are in log millilamberts)

Inducing-Field Luminance (Log B_i)	Exp. I		Exp. II		Exp. III	
	Method of Limits	Method of Con. Stim.	Method of Adjust.	Method of Con. Stim.	Method of Limits	Method of Adjust.
	.60*	.60*	.60*	.60*	.60*	.60*
−1.51	.70	.66	.65	.64	.59	.42
−0.43	.66	.72	.41	.56	.52	.72
0.06	.75	.63	.32	.50	.52	.67
0.60	.69	.44	.36	.43	.55	.62
1.14	.40	.26	.20	.32	.29	.21
1.63	.13	−.08	.07	.01	−.04	−.09
2.16	−.07	−.27	−.11	−.18	−.27	−.44
2.71	−.32	−.41	−.35	−.30	−.60	−.61

* Test-field luminance.

these complaints did not appear under the two other methods. There were no differences between the methods of limits and adjustment in this respect. In terms of time consumption and fatigue, they were equally efficient.

With regard to the actual settings under each method, these are presented in Table 1 and Fig. 1. Since there were no systematic differences between Ss within each experiment, their data were averaged.

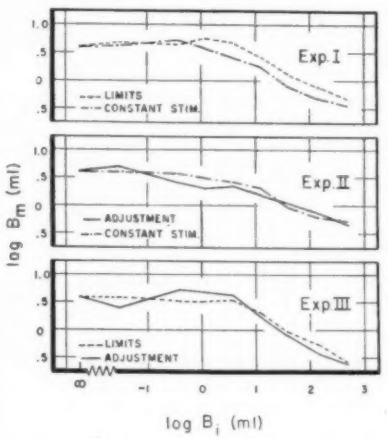

FIG. 1. Test-field apparent brightness (B_m) as a function of inducing-field luminance (B_i) as investigated under three different psychophysical methods. The test-field luminance is .60 log millilamberts for each curve.

The general shape of the function in Fig. 1 is the same for all methods in all experiments. Test-field apparent brightness is depressed slightly, if at all, for inducing-field luminances less than that of the test field. For inducing-field luminances greater than that of the test field, test-field apparent brightness is depressed roughly in direct proportion to the luminance of the inducing field. This is in agreement with previous experiments (**1, 4**).

Although the general shape of the function remains the same, the extent of the depression of the test-field apparent brightness is slightly different for the two methods in Exp. I, i.e., the test-field apparent brightness is somewhat more depressed under the method of constant stimuli than under the method of limits. It should be pointed out, however, that in Exp. II and III, both limits and constant stimuli give the same results as ad-

justment, so that the discrepancy between the curves in Exp. I cannot clearly be attributed to differences in the methods themselves.

A comparison can be made between the variances of all judgments under the methods of adjustment and limits. Constant stimuli is not included in this comparison since, according to Woodworth (**6,** p. 427), variability computed in this method is not directly comparable to that computed from the other two methods. This comparison reveals no difference between variances in the methods of adjustment and limits. The median variance in each method is .02 log ml.

Discussion

The results indicate that the inhibitive effect seen in contrast experiments (**1, 2, 4, 5**) is not peculiar to a particular type of methodology used in its investigation. The characteristic contrast effect appeared in all three methods used in the present paper.

It would seem, however, that the method of constant stimuli is less useful in studying contrast because of the excessive time and fatigue involved in procedure. In this respect, the methods of adjustment and limits are decidedly more acceptable since an entire curve can be obtained easily, without fatiguing *S* or *E*, in one experimental session.

With regard to reliability, both adjustment and limits are equally acceptable in investigating contrast.

Summary

Foveal simultaneous brightness contrast was investigated with respect to the psychophysical methods generally used in its investigation. The apparent brightness of a 33′ test-field square was studied as a function of the luminance of a 33′ adjacent inducing-field square. A binocular matching technique was used to measure the test-field apparent brightness. Three psychophysical methods of adjustment, limits, and constant stimuli were investigated.

The results indicate no difference in the characteristic shape of the contrast curve for the different methods. The methods of adjustment and limits were equally good in terms of reliability and convenience, whereas the method of constant stimuli was much more fatiguing for both *S* and *E* and took about twice as much time for the same amount of data obtained with the other two methods.

References

1. Diamond, A. L. Foveal simultaneous brightness contrast as a function of inducing- and test-field luminances. *J. exp. Psychol.*, 1953, **45,** 304–314.
2. Fry, G. A., & Alpern, M. The effect of a peripheral glare source upon the apparent brightness of an object. *J. Opt. Soc. Amer.*, 1953, **43,** 189–195.
3. Graham, C. H. Behavior, perception, and the psychophysical methods. *Psychol. Rev.*, 1950, **57,** 108–120.
4. Leibowitz, H., Mote, F. A., & Thurlow, W. R. Simultaneous contrast as a function of separation between test and inducing fields. *J. exp. Psychol.*, 1953, **46,** 453–456.
5. Schouten, J. F., & Ornstein, L. S. Measurements on direct and indirect adaptation by means of a binocular method. *J. Opt. Soc. Amer.*, 1939, **29,** 168–182.
6. Woodworth, R. S. *Experimental psychology.* New York: Holt, 1936.

(Received October 15, 1954)

Journal of Experimental Psychology
Vol. 50, No. 3, 1955

PSYCHOMOTOR PERFORMANCE AS A FUNCTION OF AMOUNT OF TRAINING AND STRESS

ALFRED CASTANEDA AND DAVID S. PALERMO[1]

Iowa Child Welfare Research Station,
State University of Iowa

A comparatively recent development is the work concerned with performance in complex learning tasks as a function of stress. A recent review of this work (**4**) indicates that while a variety of methods have been used to vary stress, e.g., failure, excessive pacing, threat, etc., the common assumption is that they affect the level of motivation. The findings, with some exceptions, indicate that performance under stress is either facilitated or impaired. The theoretical and practical importance of studies attempting to identify those variables determining the nature of the effect is generally recognized.

Another recent development in the area of motivation is the work of Spence and his colleagues (**2, 8, 10, 13**) concerned with the integration of anxiety within Hull's (**3**) framework. On the assumption that anxiety has motivational (drive) properties, certain implications have been derived regarding performance as a function of variations in the motivational level. While this work has dealt more specifically with anxiety, it has been indicated (**8**) that the more general concern has been in the isolation of some of the factors which determine the quality of performance as a function of variations in the motivational level. On the assumptions that habit and drive jointly determine response strength, and that the total effective drive level (D) then operating combines multiplicatively with each of those habit tendencies aroused in a given situation, implications have been derived concerning the quality of performance in complex learning tasks, i.e., those involving strong competing responses. If the habit strengths of the competing incorrect responses are greater in relation to the correct ones, impairment of performance should be expected with an increase in drive. However, if the correct responses are dominant, i.e., their strength is greater in relation to the incorrect ones, facilitation should be expected. Furthermore, with an increase in drive, the degree of impairment or facilitation should be positively related to the relative strength of the dominant habit at the time of the increase. These implications have been confirmed in their now familiar series of studies (**2, 7, 8, 9, 10, 11, 12, 13**) in which the Taylor Anxiety Scale has been employed as the empirical variable assumed to reflect differences in the drive level of human *S*s.

On the assumption that stress, varied by emphasizing to *S* the need for speed on the task, affects the level of drive (**1, 6**), the present study is concerned with the effects of stress on the quality of performance in a complex learning task. The task is designed to permit its analysis into specific S-R components and to permit the specification of the strength of the habit mediating each component as well as whether the habit is correct or

[1] The authors wish to express their gratitude to Mr. D. S. Stewart, Elementary School Curriculum Coordinator, Iowa City, and Mr. R. Hedeman, Superintendent of Schools, West Branch, for their cooperation and assistance.

incorrect. This is accomplished in a paired-associate type psychomotor task on which varying amounts of preliminary practice are given whereupon the task is to re-pair under stress certain of the stimulus and response elements while the others are left unchanged. According to the assumptions just outlined, the effects of stress (*D*) on the quality of performance during the relearning depend on the type of S-R pair, i.e., on whether the pair has been changed or left unchanged. If performance to the changed pairs is considered, the effect should be deleterious. If performance to the unchanged pairs is considered, the effect should be facilitation (fewer errors). Furthermore, the degree of this impairment or facilitation should be a positive function of the relative strength of the habit mediating a given S-R connection.

Method

Subjects and design.—From four different schools, 108 fifth-grade *Ss* were assigned randomly, in equal numbers, to one of four experimental groups. The *Ss* in Groups I (weak habit, nonstress) and II (weak habit, stress) were given 25 training trials on the preliminary task; those in Groups III (strong habit, nonstress) and IV (strong habit, stress) 50 training trials. All groups were given 50 relearning trials at which time Groups II and IV were administered the stress condition. Ignoring the factor of type of S-R pair, the design is a 2 × 2 factorial in which the interaction between habit strength and stress is of primary concern, with the expectation of greater differences in performance, to both types of S-R pairs, between Groups III and IV as compared to Groups I and II.

Apparatus and procedure.—The apparatus consisted of a rectangularly shaped box, 9½ × 18 × 9 in. painted flat black. Centered 3 in. above the response panel was a 1-in. diameter exposure aperture of flashed opal glass. The stimuli were five 6.3-v. pilot lamps colored green, amber, red, blue, and light red. The coloring was done by means of a General Cement Mfg. Co. Dial Lite Coloring Kit. With the exception of the red which received three dips in the coloring solution, all other lights received two dips. The lights were housed in a cluster in a No. 2½ tin can which was coated with aluminum paint to increase reflection. The housing was fastened inside of the box with the open end flush to the exposure aperture. All controls used by *E* were set to the back of the apparatus. By means of a holding circuit in which *S*'s response broke the circuit *E* could actuate any one of the five lights and set any single push button to break the circuit by appropriate adjustments of two rotary selector switches. Hence *S* could deactivate the light only by selecting the button designated correct for that light. Self-corrected trials were employed with a trial defined as a single presentation of a stimulus. Of the five different light-button pairs used, three were changed for the relearning trials while two remained the same as in the training trials. With the restriction that no single light appeared twice in succession, the order of appearance of the lights was determined randomly within blocks of five trials for the first 25 trials. Where the procedure called for 50 trials, this sequence was repeated for the next 25 trials.

The instructions were designed to indicate to *S* that he was required to associate which push button turned off a given colored light. For the relearning *S* was informed that "some" of the light-button pairs had been changed and that "some" remained the same as in the training trials. For *Ss* in the stress condition, additional instructions were given to the effect that speed was the measure of their learning ability, and a Standard Electric timer was connected into the apparatus and placed so that its face was clearly visible to him. The circuit was such that the timer was actuated simultaneously with the onset of a light and remained on until *S* made the correct response which automatically broke the circuit to both the timer and the light. Incorrect responses affected neither the timer nor the light. For all groups the relearning trials were administered upon completion of the training trials. Every response made to each light, and the order in which it was made, was recorded for each *S*.

Results

Error scores of the four groups for the first 25 training trials were subjected to a simple analysis of variance to test the hypothesis that they were random samples from the same population. For 3 and 104 *df*, the obtained *F* of 1.12 did not permit rejection of the null hypothesis.

According to the present assumptions, if the dominant habit is incor-

TABLE 1

ERRORS TO CHANGED AND UNCHANGED S–R PAIRS ON RELEARNING TRIALS 1–20

Level of Habit Strength	S–R Pair											
	Unchanged						Changed					
	Nonstress			Stress			Nonstress			Stress		
	Group	Mean	*SD*	Group	Mean	*SD*	Group	Mean	*SD*	Group	Mean	*SD*
Weak	I	5.41	3.16	II	5.96	4.78	I	15.22	6.56	II	17.52	6.01
Strong	III	5.52	3.01	IV[1]	4.70	3.41	III	11.52	4.89	IV	19.96	6.44

rect, the consequences of an increase in drive should be deleterious, but facilitation would be expected if the dominant habit is correct. Augmentation of either of these effects would be expected with increases in the strength of the dominant habit over those competing. Table 1 presents the data relevant to these expectations based on the first 20 relearning trials.[2] It can be seen that in the case of the changed S-R pairs (dominant habit incorrect) the general effect of an increase in drive is deleterious with the tendency for this effect to be augmented by increases in the strength of the habit. However, in the case of the unchanged S-R pairs (dominant habit correct) the effect of an increase in drive for the strong habit group is in the direction of facilitation.

A summary of an analysis of variance (Lindquist Type III [5]) based on these data is presented in Table 2. It can be seen that the over-all differences due to stress are significant as are the differences due to the type of S-R pair, this latter undoubtedly reflecting differences due to the relatively fewer number of presentations of the unchanged pairs over the first 20 trials. Of particular theoretical interest, however, is the significant interaction between stress, amount of training and the type of S-R pair. This may be interpreted as indicating that the effects of stress with varying amounts of training are not the same for the two types of S-R pairs. This is in accord with the data presented in Table 1, i.e., an increase or decrease in errors due to stress depends on the type of S-R pair considered.

[2] Although the procedure called for 50 relearning trials, justification lies within the present theory for the expectation that the differences expected would be maximal in the early trials. Analysis based on the 50 trials disclosed that a number of *Ss* in the stress condition had learned the task. Analysis based on the first 10 trials revealed that not only were all differences in the expected direction but that, in comparison to the differences based on 20 trials they were in most cases appreciably greater. However, the scores based on the first 10 trials were too variable to be reliable.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF ERROR SCORES BASED ON TRIALS 1–20

Source	*df*	*MS*	*F*	*P*
Training	1	19.56	—	
Stress	1	370.78	10.96	<.005
Training × stress	1	77.04	2.28	<.20
Error (b)	104	33.82		
S-R Type	1	6133.34	333.44	<.001
S-R Type × training	1	.04	—	
S-R Type × stress	1	408.38	22.20	<.001
S-R Type × training × stress	1	190.28	10.37	<.005
Error(w)	104	18.39		

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE OF ERRORS TO CHANGED S-R PAIRS ON RELEARNING TRIALS 1–20

Source	*df*	*MS*	*F*	p
Training	1	6.08	—	
Stress	1	778.71	22.31	<.001
Training × stress	1	529.77	15.18	<.001
Error (w)	104	34.91		

Since all *F*'s based on scores to the unchanged pairs were less than unity, indicating the lack of reliability of the differences shown in this category in Table 1, only the results of the analysis of variance for the changed pairs are presented in Table 3. These results indicate that the main effect of stress was significant while that of training was not. However, it is the significant interaction between these two variables which is of theoretical importance. Further analysis indicated that this interaction was due mainly to the differences between Groups III and IV ($P < .001$, $F = 28.35$, $df = 1$ and 52) as compared to the difference between Groups I and II ($P < .20$, $F = 1.73$, $df = 1$ and 52). These results are in accord with the assumption that these two variables, habit strength and drive, combine multiplicatively in the determination of response strength. Other analyses indicated that Group III performed better than Group I ($P < .05$, $F = 5.33$, $df = 1$ and 52) and that Group II performed better, but not significantly better, than Group IV ($P < .20$, $F = 2.00$, $df = 1$ and 52).

DISCUSSION

The results of this study tend to be in agreement with the implications of the assumption relating habit strength and drive to response strength. With one exception the differences, while not all significant, were in the expected direction.

Though these results tend to be in accord with the theory, a number of alternative hypotheses should be considered. Since the stress condition required *S* to alter his pace, it could be assumed that its main effect was to disrupt a pacing set with a resulting increase in the number of errors. A second explanation could be based on the assumption that the stress condition served as a distracting element serving to impair performance as reflected in increases in the number of errors. Still, a third alternative explanation might be based on the assumption that the stress instructions which emphasized speed may have resulted in a set to respond rapidly with a disregard for accuracy. On each of these assumptions, however, it would seem plausible to expect impairment of performance to both the changed and unchanged S-R pairs. It is clear from Table 1 that the effects of stress are neither uniformly impairing nor facilitating, but depend on whether the dominant habit is correct or incorrect.

In conclusion, these results accord some support to the contention that stress, as manipulated in this study, exhibits some of the properties of drive. Additionally, support is given to the specific manner in which the variables of habit strength and drive are assumed to be related in the determination of response strength. The particular manner in which the dominant habit was varied as to whether it was correct or incorrect was such that, on the basis of these assumptions, both facilitation and impairment on the same task could be expected as a consequence of an increase in the drive level. To the extent that these expectations were realized, support is given to the notion that predictions regarding the quality of performance as a function of variations in the motivational level (*D*) requires knowledge of the strengths of the specific habits affected by such variations and whether the dominant habit is correct or incorrect.

SUMMARY

Performance as a function of the amount of training and stress was studied on the assump-

tion that stress, under the conditions of this study, affects the level of drive. Impairment of performance was predicted as a result of increases in drive where the dominant habit was incorrect and facilitation where it was correct. Also, it was predicted that the degree of impairment or facilitation under such conditions would be a positive function of the relative strength of the dominant habit.

Of four groups of 27 *Ss* each, two were given 25 training trials on a preliminary task, and two were given 50 trials. All groups were given 50 relearning trials in which three of the stimulus and response elements had to be re-paired while the remaining two were left unchanged. One group from each of the two training groups relearned under stress. Stress was found to increase significantly the number of errors to the changed S-R pairs, with this increase being more pronounced with increased training on these pairs. A tendency, though unreliable, was found for stress to decrease the number of errors to the two unchanged pairs. On the basis of these results, it was concluded that predictions regarding the effects on performance of variations in the drive level requires knowledge of the strengths of the specific habits involved and whether the dominant habit is correct or incorrect.

References

1. Alper, T. G. Memory for completed and incompleted tasks as a function of personality; an analysis of group data. *J. abnorm. soc. Psychol.*, 1946, **41**, 403–421.
2. Farber, I. E., & Spence, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1956, **45**, 120–125.
3. Hull, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
4. Lazarus, R. S., Deese, J., & Osler, S. F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.
5. Lindquist, E. F. *Design and analysis of experiments in psychology and education*. New York: Houghton Mifflin, 1953.
6. Lindsley, D. B. *A study of performance under speed stress*. OSRD, 1945, Publ. Bd. No. 18381. Washington, D. C.: U. S. Dept. of Commerce, 1946.
7. Montague, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.
8. Spence, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. In *Kentucky Symposium, Learning theory, personality theory and clinical research*. New York: Wiley, 1954, Ch. 1.
9. Spence, K. W., & Farber, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.
10. Spence, K. W., & Farber, I. E. The relation of anxiety to differential eyelid conditioning. *J. exp. Psychol.*, 1954, **47**, 127–134.
11. Spence, K. W., & Taylor, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
12. Taylor, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 91–92.
13. Taylor, J. A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.

(Received October 8, 1954)

Journal of Experimental Psychology
Vol. 50, No. 3, 1955

AMOUNT OF PRETRAINING AS A FACTOR IN STIMULUS PREDIFFERENTIATION AND PERFORMANCE SET [1]

JOAN H. CANTOR

State University of Iowa [2]

Recent research has indicated a growing interest in the effects of various kinds of preliminary verbal training upon the later acquisition of skill in performing a motor task. A number of investigators (**1, 2, 3, 5, 6, 7, 8, 10**) have studied the effects of pretraining in which *S* gains knowledge about the motor task stimuli. This pretraining has generally consisted of verbal paired-associates learning in which the stimuli are the same as, or substitutes for, those of the motor task, while the verbal responses are unrelated to the required motor responses. Increased discriminability among the motor task stimuli, as a result of the pretraining, has been referred to either as stimulus predifferentiation or acquired distinctiveness of cues. Several studies using stimulus predifferentiation training have demonstrated positive transfer to the motor task (**1, 2, 5, 7, 8**), but two investigators have found evidence of an interfering effect (**1, 10**).

Superior motor performance has also been shown to be a result of the transfer of nonspecific factors following paired-associates learning in which both the stimuli and responses are unrelated to those of the motor task (**6**). Because of these effects of "performance set," an "irrelevant" [3] pretraining group is necessary as a control in the investigation of stimulus predifferentiation.

In a systematic investigation of stimulus predifferentiation and performance set, it is important to determine the relationship between amount of pretraining and subsequent motor performance.

Gagné and Baker (**7**) gave three amounts of relevant pretraining prior to a motor task requiring the operation of a particular toggle switch for each of four colored stimulus lights. The group receiving the largest amount of pretraining performed significantly better on the motor task than did a group receiving none. However, this facilitation could be attributed to the effects of performance set as well as to stimulus predifferentiation.

Baldwin (**1**) gave two amounts of relevant pretraining prior to performance on a simple discrimination task. For the larger amount, *S*s given relevant pretraining made significantly more correct responses on the motor task than did a group given the same amount of irrelevant pretraining.

The purpose of the present study was to investigate the effects of three amounts of relevant pretraining upon the subsequent performance of a perceptual-motor task. The study was also designed to obtain information concerning the influence of the same

[1] This article is based on a dissertation submitted to the Graduate College of the State University of Iowa in partial fulfillment of the requirements for the Ph.D. degree. The writer is indebted to Dr. Don Lewis for advice and assistance throughout the course of the investigation.

[2] Now at George Peabody College for Teachers, Nashville, Tenn.

[3] The terminology used is the same as that suggested by McAllister (**10**). The word "irrelevant" refers to the specific stimulus and response elements in the paired-associates task and does not imply that all aspects of the pretraining task are unrelated to the motor task. Stimulus predifferentiation training has been called "relevant stimulus" pretraining and is here abbreviated to "relevant" pretraining.

amounts of irrelevant pretraining upon motor performance.

The Star Discrimeter provided the criterion task in which *S* learned a particular motor response for each of six similar colored lights. The *S* moved a wobble stick into one of six horizontal channels when the corresponding color was presented. During relevant pretraining, *S*s learned to associate a letter with each of these colors, and those given irrelevant pretraining associated the same letters with a different set of colors.

One of the earliest explicit statements of the mechanism underlying stimulus predifferentiation is that of Miller and Dollard (**11**, p. 73). Applying their formulation to the present situation, it is assumed that the verbal responses learned during relevant pretraining transfer to the motor task. Presumably, these verbal responses have stimulus components, and each light stimulus therefore has a constant additional cue which serves to distinguish it from the other stimuli in the group. The result is decreased generalization among the colors and faster learning of the motor task. With increased amounts of relevant pretraining, the probability that the correct verbal responses will transfer to the motor task should increase, resulting in increased discrimination and thus greater facilitation.

Performance set involves the transfer of less specific responses and includes such factors as warm-up and learning-to-learn. Therefore, *S*s given irrelevant pretraining may show motor performance superior to that of *S*s given none.

On the basis of theory and available data, certain predictions were made at the outset of the experiment: (*a*) Relevant pretraining would result in positive transfer to the motor task for all three amounts of pretraining. Specifically, it was believed that the relevant groups would show performance superior to that of the irrelevant groups. (*b*) The amount of positive transfer to the motor task would increase with increasing amounts of relevant pretraining. (*c*) Irrelevant pretraining would result in positive transfer to the motor task for all three amounts of pretraining. All irrelevant groups were expected to show better motor performance than that of a no pretraining group.[4]

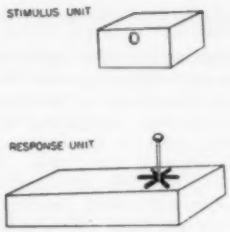

FIG. 1. Basic elements of the Star Discrimeter.

METHOD

Apparatus and procedure.—The motor task was provided by the Star Discrimeter, the Iowa version of a device described by Duncan (**4**). As seen in Fig. 1, six channels are cut into a horizontal steel plate so that they radiate out at 60° angles from a center opening. A steel wobble stick protrudes from the center opening and may be moved freely into any one of the channels. Mounted at about eye level behind and slightly to the left of the wobble stick is a black stimulus box. A circular piece of opal glass is placed in the opening near the top edge of the box. Any one of six colors may be diffused onto the glass from inside the box. A 50-point stepping relay is used to present a continuous random sequence of the six colors.

The colors consisted of red, orange, and yellow hues. Although it was not possible to obtain a completely continuous series of hues, each color

[4] No prediction was made regarding the relative effects of larger amounts of irrelevant pretraining upon performance set. The writer knows of no empirical evidence concerning these effects in a situation using pretraining with a motor task.

was quite similar to at least one other color in the series.

The *S* stood in front of the apparatus, holding the wobble stick with the preferred hand. The task was to learn which of the six colors was associated with each channel and to move the wobble stick into as many correct channels as possible during each trial period. The *S* continued moving the stick into channels until the correct channel was entered, resulting in the appearance of a new color. Microswitches placed in the channels activated correct response, shallow, and deep error counters.

Pretraining was given in the same location as motor training. The colors were presented in the manner described above, and, for the relevant groups, were identical to those of the motor task. For the irrelevant groups, a series of blue-green hues was used. The two sets of colors had been roughly equated for discriminability during preliminary work.

An automatic slide projector placed directly in front of the stimulus box was used to present the letters. A series of timers provided automatic synchronization of the stepping relay on the Star and the slide projector. Each color in the sequence was presented for 4 sec. A blank slide was presented during the first 2 sec. followed by a slide containing the letter associated with that color during the last 2 sec. The letter appeared directly underneath the color. The procedure was therefore exactly like paired-associates learning. A random sequence of 36 colors was used for each block of stimuli, every color being presented six times.

The group without pretraining received instructions and immediately began practice on the Star Discrimeter. The procedures were the same for all pretraining groups except that the relevant groups associated letters with the red-yellow hues, and the irrelevant groups with the blue-green hues. The instructions explained that *S* was to learn to discriminate similar colors by associating a particular letter with each of the colors and pronouncing the correct letter before it was shown. The pretraining groups were given 2, 4, or 12 blocks of stimulus presentations with the appropriate colors. Two-minute rests were given after every three blocks. At the end of pretraining, *S* left the room for a 2-min. rest while *E* set up the motor task. The projector was rolled away and the wobble stick was uncovered for the first time. For the irrelevant groups, the colors were changed. There were 40 trials on the Star Discrimeter, each trial lasting 20 sec., with 10-sec. intertrial intervals. A 2-min. rest was given after 20 trials.

Subjects.—Women volunteers from a course in introductory psychology were used as *S*s, and 20 women were randomly assigned to each of the seven groups. All *S*s were required to read the numbers correctly on Cards 1–9 and 12–13 of the Ishihara color test. Seventeen *S*s were discarded because of color blindness, apparatus breakdown, or failure to follow instructions.

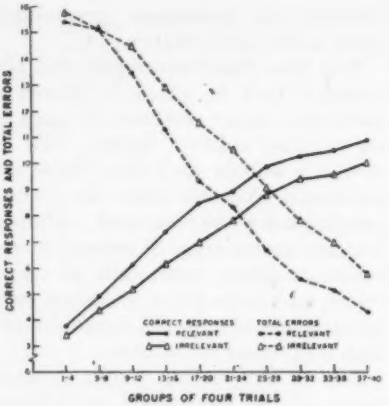

FIG. 2. Means per trial in blocks of four trials on the Star Discrimeter for the combined relevant group and the combined irrelevant group.

RESULTS

Pretraining performance.—For each amount of pretraining, no significant difference was found between the relevant and irrelevant groups in either total number of correct responses or in the number of correct responses on the last pretraining block.[5] Therefore, differences in motor performance could not reasonably be attributed to possible differences in the difficulty of the pretraining task.

Motor performance.—Means per trial of correct responses and of total errors for the three relevant groups combined and the three irrelevant groups combined are plotted against blocks of four trials in Fig. 2. (Total error scores were the sum of deep and

[5] Whenever possible, the assumption of homogeneity of variance was tested and was found to be tenable in all cases. A 5% coefficient of risk was adopted for all statistical analyses.

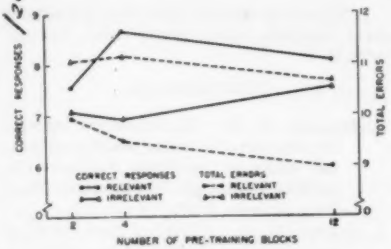

FIG. 3. General means per trial over Trials 1–40 on the Star Discrimeter for each pretraining group.

shallow errors.) General means per trial over Trials 1–40 are presented in Fig. 3.

The correct response and error data for all 40 trials were analyzed for the six pretraining groups according to Lindquist's (9) Type III design, the variables being type of pretraining, amount of pretraining, and trials. The results were essentially alike for the two measures of performance. The difference between the relevant and irrelevant groups was significant for both correct responses ($F = 5.96$, $df = 1, 114$) and total errors ($F = 4.91$, $df = 1, 114$). No significant differences were found for amount of pretraining, nor were the interactions between type and amount of pretraining significant. The trials effect was highly significant, as indicated by the curves in Fig. 2. The interaction between type of pretraining and trials was significant for both correct responses ($F = 1.79, df = 39, 4446$) and total errors ($F = 1.94, df = 39, 4446$), showing that the trends in Fig. 2 are not parallel over all 40 trials.

No significant differences in motor performance were found among the irrelevant and no pretraining groups.

DISCUSSION

The data support the first prediction that all three amounts of relevant pretraining would have a facilitative effect. As seen in Fig. 2, the superiority of the combined relevant group was greatest during the middle trials, leading to the significant interaction between trials and type of pretraining. The point in learning at which facilitation should begin to be evidenced is a difficult one to determine. Decreased generalization among the stimuli is not expected to be facilitating until at least some of the correct associative connections are established in the motor task. Before any correct associations are formed, the identification of the individual colors should not result in more correct responses or fewer errors. However, as soon as some learning occurs, increased discriminability of the stimuli should allow for faster elimination of incorrect responses. The present data showed an increase in the superiority of the relevant group followed by a decrease as the performance asymptote was approached. Presumably, the differences would have completely disappeared had practice been continued.

Contrary to the second prediction, the amount of facilitation did not increase with increased amounts of relevant pretraining. The expectation was that the relevant-irrelevant trend lines (see Fig. 3) for correct responses and for total errors would diverge, the effect being reflected by significant interactions between type and amount of pretraining. However, the hypotheses of parallel trends could not be rejected. The facilitation for four pretraining blocks was considerably greater than that for two blocks, as predicted. However, the largest amount of relevant pretraining resulted in relatively little positive transfer.

Although it now appears clear that stimulus predifferentiation can be shown to be facilitative, the variables determining the amount of facilitation are difficult to specify. Since the facilitation depends upon the transfer to the motor task of responses learned during pretraining, amount of pretraining should be an important variable. The results suggest that the functional relationship between amount of pretraining and amount

of facilitation is not a simple one, and that other relevant variables must be identified before the nature of the function can be determined.

No verification of the third prediction was found. The motor performance of the irrelevant and no pretraining groups was essentially the same. Apparently any transfer of nonspecific factors to the motor task which occurred did not facilitate motor performance. Further systematic studies of the relevant variables are needed to explain the differences in experimental findings regarding performance set.

Summary

The purpose of this study was to investigate the effects of three amounts of verbal pretraining upon performance of a perceptual-motor task. In the motor task provided by the Star Discrimeter, *Ss* learned to move a wobble stick into channels corresponding to six colored lights consisting of red-yellow hues. Three groups of 20 women received either 2, 4, or 12 blocks of relevant pretraining in which they learned to associate a letter with each of the six colors. Three irrelevant pretraining groups were given corresponding amounts of practice in associating letters with a blue-green series of colors. An additional group had no pretraining prior to the motor task. All *Ss* were given 40 trials on the Star Discrimeter. The performance measures were correct responses and total errors per 20-sec. trial.

The predictions made and the results obtained were as follows:

1. The data supported the prediction that all three amounts of relevant pretraining would result in positive transfer to the motor task. The performance of the combined relevant groups was reliably superior to that of the combined irrelevant groups. This facilitation was attributed to decreased generalization among the stimuli of the motor task.

2. The prediction that amount of facilitation resulting from stimulus predifferentiation would increase with increasing amounts of pretraining was not verified.

3. The prediction that irrelevant pretraining would facilitate motor performance was not verified.

References

1. Baldwin, R. D. Discrimination learning as a function of stimulus predifferentiation and mediated association training. Unpublished doctor's dissertation, State Univ. of Iowa, 1954.
2. Cantor, G. N. Effects of three types of pretraining on discrimination learning in preschool children. *J. exp., Psychol.*, 1955, **49**, 339–342.
3. Cantor, J. H. The effects of pre-training on the performance of a simplified Mashburn task. Unpublished master's thesis, State Univ. of Iowa, 1953.
4. Duncan, C. P. Transfer in motor learning as a function of degree of first-task learning and inter-task similarity. *J. exp. Psychol.*, 1953, **45**, 1–11.
5. Dysinger, D. W. An investigation of stimulus pre-differentiation in a choice discrimination problem. Unpublished doctor's dissertation, State Univ. of Iowa, 1951.
6. Farber, I. E., & Murfin, F. L. Performance set as a factor in transfer of training. Paper read at Midwest. Psychol. Ass., Chicago, April, 1951.
7. Gagné, R. M., & Baker, K. E. Stimulus pre-differentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439–451.
8. Gerjuoy, I. R. Discrimination learning as a function of the similarity of the stimulus names. Unpublished doctor's dissertation, State Univ. of Iowa, 1953.
9. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
10. McAllister, D. E. The effects of various kinds of relevant verbal pretraining on subsequent motor performance. *J. exp. Psychol.*, 1953, **46**, 329–336.
11. Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univ. Press, 1941.

(Received October 20, 1954)

Journal of Experimental Psychology
Vol. 50, No. 3, 1955

TRANSPOSITION IN THE WHITE RAT AS A FUNCTION OF STIMULUS COMPARISON[1]

ROBERT THOMPSON

Louisiana State University

Recent work in discriminative learning has revealed a number of difficulties for the two main interpretations of transposition behavior. The theory held by Köhler (**4**) and Lashley (**5**) which assumes that *S* responds to the relationship existing between the stimuli suffers from the findings that this cue can be experimentally abolished without disturbing the discriminative habit (**2, 3, 11**). In these studies, the habit was apparently maintained on the basis of the absolute properties of the stimulus objects. Similarly, the alternate theory of Spence (**9, 10**) proposing that *S* responds to the specific aspects of the stimulus situation receives little support from the results obtained by Nissen, Levinson, and Nickols (**7**), and by Saldanha and Bitterman (**8**). These investigators showed that relational cues seem to operate in the mastery of certain discrimination problems.

In the light of this conflicting evidence, it is not unreasonable to assume that *S* learns to respond to both the relative and absolute properties of the stimulus situation. Nissen *et al.* (**7**) came to a similar conclusion in explaining the performance of chimpanzees on a complex discrimination.[2] This interpretation is also applicable to transposition behavior, e.g., the degree of transposition being a function of the interaction between relational and specific learning. Previous studies have been primarily interested, however, in the effect of only one variable on transposition performance, and such data are inadequate to evaluate the necessity of positing response to two different cues.

The aim of the present experiment was to study the effects of three variables on transposition performance. The degree of comparison of training stimuli appears to be an important variable as shown by Baker and Lawrence (**1**). These investigators, using a Grice apparatus, failed to demonstrate transposition in rats whose training did not afford them direct comparison of the training stimuli. Although these results support a relational learning theory, the authors admit that they are not crucial because of the difficulties inherent in the apparatus employed. Weiss and Stevenson (**12**) recently investigated the effect of delayed testing on transposition in adult human *Ss*. They found that increasing amounts of time between attainment of criterion and the test tended to increase the degree of transposition. The third variable about which the present study was concerned dealt with the degree of difficulty of the training problem, a

[1] This paper represents a portion of a dissertation submitted to the Faculty of the Graduate School of the University of Texas in partial fulfillment of the requirements for the Ph.D. degree. The author is indebted to Dr. H. C. Blodgett for advice during the course of this study.

[2] The majority of relational theorists have mentioned the existence of response to the absolute properties (**4, 5**), but relegate it to an insignificant role in view of the fact that memory of relations shows more permanence and stability than memory of absolute properties.

variable which has not received systematic study in animals.

Method

Subjects and Apparatus

Sixty-four experimentally naive albino rats, 32 males and 32 females, were used as *Ss*. Their ages ranged from 60 to 90 days at the beginning of the experiment. All *Ss* were handled for approximately three weeks prior to the beginning of experimental training. During the last week, each *S* was placed on a 23-hr. feeding schedule.

A conventional two-window Lashley jumping apparatus was employed, each window measuring 4.5 in. square. Two feet below the windows was a cloth net into which the *S* fell upon making an incorrect response. Behind the windows was a feeding platform where a constant supply of wet mash was provided throughout the experiment. The entire apparatus was painted a flat black.

Procedure

Preliminary training.—The *Ss* were fed on the food platform for a period of one week, during which time they became accustomed to the apparatus through exploration. At the end of this period *Ss* were trained for five days to jump through open windows. At first the jumping platform was moved in so that the gap between it and the windows was small, enabling *S* to walk through the open windows. Gradually the gap was increased to a maximal distance of 9 in. Later all *Ss* were trained to jump through vertical black- and white-striped cards (striations .5 in. in width) which were placed unlocked against the windows. During the entire preliminary training period *Ss* were given guided trials, if necessary, to prevent the development of a consistent preference for either of the two windows.

Training pairs.—Following preliminary training, *Ss* were divided at random into two groups of 32 *Ss* each. Each group was trained concurrently on two separate pairs of stimulus cards, the members of one pair differing only in reflectance, the members of the second pair differing only in size of circle outlines. One group (Comparison Group) received training on the two pairs which afforded them direct comparison between cards in the same dimension. That is, the members of the brightness pair or the members of the size pair appeared together in the two windows. The second group (Noncomparison Group) was trained on the two pairs so that direct comparison of cards in the same dimension was impossible. In this case, appearing in the two windows were either one member of the size pair and one member of the brightness pair, or the other two members of each pair. The technique employed was the same as that used by Saldanha and Bitterman (8). Table 1 illustrates the training pairs for the Comparison and Noncomparison Groups.

TABLE 1

Arrangement of Pairs for the Comparison and Noncomparison Groups *

Comparison Group		Noncomparison Group	
+	−	+	−
Light gray	Dark gray	Light gray	Large circle
+	−	+	−
Small circle	Large circle	Small circle	Dark gray

* Circles were white outlines ¼ in. wide painted on black backgrounds.

TABLE 2

Stimulus Values for the Cards Used in the Brightness and Size Discriminations

Problem	Brightness Discrimination (Per Cent Reflectance)		Size Discrimination (Outside Diameter, In.)	
	L +	D −	c +	C −
Easy	68	25	2.50	4.25
Difficult	68	43	2.50	3.50

Half of the *Ss* in each main group received training on an "easy" problem, the other half received training on a "difficult" problem. The problems were made to vary in difficulty by decreasing or increasing the intrapair difference in reflectance for the brightness pair and in diameter for the size pair. Table 2 gives the stimulus values in reflectance and the sizes of circle diameter for the two pairs involved in the easy and difficult problems.

A total of 12 trials was given each day, 6 trials on each pair. The pairs were presented in a prearranged random order. The positive card of each pair appeared on the right for three trials and on the left for the other three trials. A response to the positive card of either pair admitted *S* to the rear platform, whereas a response to the negative card, which was locked, caused *S* to fall into the net. The method of correction was employed throughout, each trial therefore terminating in a correct response. If *S* failed to give a correct response in three "free" jumps, *S* was manually guided to the positive card.

Training was continued until *S* made no more than two errors in two consecutive days (22 errorless responses in 24 trials with no more than one error on each pair).

Transposition test.—Upon mastering the training pairs, the transposition test was given. Two different time intervals between the mastery of the training pairs and the first 12 trials on the transpositional stimulus cards were employed. Half of the *S*s in each subgroup were tested immediately upon reaching criterion. The remaining half were tested 24 hr. after meeting criterion.

The transposition test, which was given in the positive direction, consisted of the pairing of a lighter stimulus card (85% reflectance) with the previously positive light card (68% reflectance), and the pairing of a smaller circle (1.75 in. outside diameter) with the previously positive small circle (2.5 in. outside diameter). The positive card of each pair during the transposition test was the lighter stimulus card and the smaller circle. The stimulus cards were paired so that there would be direct comparison between cards on the same dimensions.

The experimental procedure used in the transposition test was the same as that used in the training period. The brightness pair and the size pair were learned concurrently. Twelve trials were given each day with six trials on each pair. The two pairs were presented in random order, the positive card of each pair appearing on the left for half of the trials. Training was terminated when *S* committed no more than two errors in two consecutive days.

Results

Original learning.—Before discussing the results of the transposition test, it is necessary to examine the performance of the Comparison and Noncomparison Groups on the easy and difficult problems in the original learning. Table 3 gives this information in terms of days to criterion and total errors. These data show that the manipulations employed to provide two training problems differing in difficulty were successful. For both the Comparison and Noncomparison Groups, the easy problem required significantly fewer days to master (t's = 5.36 and 5.64; P's = .001) and resulted in significantly fewer errors (t's = 5.18 and 5.09; P's = .001) than the difficult problem.

TABLE 3

Performance Scores on Training Problems for Comparison and Noncomparison Groups

Problem	*N*	Days to Criterion		Total Errors	
		Mean	*SD*	Mean	*SD*
Comparison Group					
Easy	16	6.0	1.52	49.2	15.31
Difficult	16	10.5	2.91	81.8	19.76
Noncomparison Group					
Easy	16	6.2	1.93	47.5	19.54
Difficult	16	11.5	3.16	89.0	25.98

More importantly, there were essentially no differences between the

TABLE 4

Percentage Correct Responses on First Trial and on First Day on the Brightness and Size Discriminations

Condition	*N*	First Trial		First Day	
		Size Pair	Brightness Pair	Size Pair	Brightness Pair
Group					
Comparison	32	72*	75**	70*	83***
Noncomparison	32	63	72*	62	72**
Problem					
Easy	32	66	72*	66	75**
Difficult	32	69*	75**	66	80***
Time of Testing					
Immediate	32	66	69*	64	74**
Delayed	32	69*	78***	68*	81***

* P = .05.
** P = .01.
*** P = .001.

comparison and noncomparison procedures, a result contrary to that obtained by Saldanha and Bitterman (8). The differences between the Comparison and Noncomparison Groups on the easy problem were very unreliable. On the difficult problem, the Comparison Group tended to be superior to the Noncomparison Group in terms of both days to criterion and total errors, but the resulting P values fell short of the .05 level of significance.

Transposition test.—One of the significant indices of transposition is performance on the first test trial with the new pair of stimuli. In the present study the use of two different training pairs therefore provided two such test trials—one on the size pair and one on the brightness pair. Table 4 presents the percentage of correct responses on the first test trial under each major experimental condition. The P values were calculated by the t test for proportions. It can be seen that there was significant transposition under each condition on the brightness pair. There was also significant transposition on the size pair, although the significance levels are marginal for the noncomparison procedure, the easy problem, and the immediate testing conditions.

In terms of combined performance on both pairs for the first trial, there was no significant difference in transposition between the comparison and the noncomparison conditions (P = .50). Similarly, no significant difference in the number of relative choices occurred between the easy and difficult problems (P = .70), or between the immediate and delayed testing conditions (P = .50).

Table 4 also provides the percentage of correct responses on the first day in terms of initial jumps on the six test trials with each pair. With this second index of transposition, it can be seen that a significant number of relative choices was obtained under all experimental conditions on the brightness pair. On the size pair, however, only the comparison and delayed testing conditions resulted in significant transposition, whereas the levels of significance for the other experimental conditions fell short of statistical significance.

The differences between the related conditions (comparison versus non-

TABLE 5

ANALYSES OF VARIANCE OF TOTAL ERRORS ON FRIST DAY, DAYS TO CRITERION, AND TOTAL ERRORS TO CRITERION

Source of Variation	*df*	Total Errors on First Day		Days to Criterion		Total Errors to Criterion	
		Mean Square	*F*	Mean Square	*F*	Mean Square	*F*
Modes of presentation (P)	1	51.0	7.61**	85.2	18.89***	1507.1	20.67***
Degrees of difficulty (D)	1	1.2		19.3	4.30*	302.3	4.09*
Time of testing (T)	1	16.0	2.40	4.9	1.09	74.6	1.01
P × D	1	2.2		1.6		16.3	
P × T	1	3.1		1.1		41.8	
T × D	1	28.3	4.23*	2.6		166.3	2.25
T × D × P	1	10.4	1.65	3.1		71.3	
Within	56	6.67		4.5		73.9	
Total	63						

* P = .05.
** P = .01.
*** P = .001.

comparison procedures, easy versus difficult problems, and immediate versus delayed testing conditions) in terms of combined correct responses on the first day did not approach the .05 level of significance.

Table 5 summarizes the analyses of variance for total errors on the first day, total errors to criterion, and trials to criterion. For the three analyses, the eight groups are of homogeneous variance as shown by Bartlett's test. On the basis of these measures of performance, the comparative procedure was superior to the noncomparative procedure in learning the transposition problem. The significant F's of 4.30 and 4.09 for days to criterion and total errors to criterion, respectively, show that the *S*s trained on the difficult problem performed more efficiently on the transposition pairs than the *S*s trained on the easy problem. Of particular interest is the significant interaction existing between time of testing and problem difficulty for total errors on the first day of the transposition test. Inspection of the data reveals that, for the easy problem, delayed testing shows more efficient performance than immediate testing. That is, the effectiveness of the delayed testing condition is limited to the easy problem.

Discussion

One of the principal findings of this experiment was the reliable evidence of transposition in the Noncomparison Group. The *S*s in this group exhibited a very significant degree of transposition on brightness in terms of performance on the first trial as well as in terms of performance on first day. On size, however, only a marginal degree of transposition was found. The fact that the noncomparison *S*s tended to respond relationally to the transpositional stimulus cards stands in contradistinction to the results reported by Baker and Lawrence (1). These investigators, using a Grice discrimination apparatus, found that the successively trained group (Noncomparison Group) not only failed to show transposition, but exhibited a slight tendency to respond to the absolute characteristics of the test situation.

At least two differences in procedure between the present study and that of Baker and Lawrence may be responsible for the conflicting findings. In the first place, the successive discrimination problem used during the training period required that only one of the two discriminanda appear on each trial. This one-choice situation used by Baker and Lawrence contrasts sharply with the two-choice situation employed for the noncomparison *S*s in the present study. It is quite possible, as Baker and Lawrence admit, that successively trained *S*s may acquire different receptor orientation habits in the one-choice situation than in the two-choice situation. Secondly, the test for transposition in the Baker and Lawrence study required switching the successively trained *S*s to a two-choice simultaneous situation, whereas the present study preserved the two-choice situation throughout both phases of the experiment. Thus there remains the possibility that the different receptor orientation responses, as well as the switching from the one-choice to the two-choice situation, may be responsible for the absence of transposition in the successively trained group employed by Baker and Lawrence.

The appearance of transposition exhibited under the noncomparative condition is consistent with the theory of Spence (9, 10). It might seem, however, that these data oppose a relational learning theory (4, 5), since direct comparison of the training stimuli on the same continuum was impossible. The relationship between the stimuli may have been built up under the noncomparative condition if *S* compared sensory impressions with previous traces. This interpretation would hold only when the intertrial interval is brief. The poorer transposition learning scores on the test problem for the noncomparative *S*s may be handled by both theories on the basis

of negative transfer of receptor orientation responses.

There was a tendency, although insignificant, for the *S*s trained on the difficult problem to exhibit more transposition in terms of first-trial and first-day performance than the *S*s trained on the easy problem. In terms of transposition learning measures, this difference became significant. The degree of transposition as a function of the difficulty of the training problem has not received attention by either of the two main theories. This variable should be taken into account when a generalized theory of discrimination learning and transposition is developed.

Although the present data do not provide critical evidence against either the relational or the specific stimulus theories, they can perhaps best be assimilated in a theory which acknowledges the occurrence of both absolute and relational leaning. Such a theory would suggest a line of further experimentation and would emphasize interactions among variables which have, heretofore, been neglected.

Summary

This experiment was designed to study the effects of comparison of training stimuli, the degree of difficulty of the training pairs, and the time of testing on transpositional behavior.

Sixty-four albino rats were given training on a jumping stand to discriminate between two discrete pairs of stimulus cards, the members of one pair differing only in reflectance, and the members of the other pair differing only in size of circle outline. Thirty-two *S*s received training which afforded them direct comparison of stimuli on the same dimension. The remaining 32 *S*s were trained so that direct comparison of the stimulus objects in the same dimension was impossible. Half of the *S*s in each group was trained on "easy" pairs; the other half was trained on "difficult" pairs. Test for transposition was given either immediately or delayed for 24 hr.

The following results were obtained:

1. Significant transposition was obtained under both the comparative and noncomparative conditions in terms of first trial and first jump on the six trials on the first day.

2. In terms of transposition learning scores, the Comparison Group yielded more transposition than the Noncomparison Group ($P = .01$).

3. In terms of total errors on first day, there was a significant interaction between degree of difficulty and time of testing ($P = .05$). Only on the easy problem was delayed testing superior to immediate testing.

4. The difficult problem yielded more transposition than the easy problem with respect to days to criterion ($P = .05$) and total errors to criterion ($P = .05$) in transposition learning.

References

1. Baker, R. A., & Lawrence, D. H. The differential effects of simultaneous and successive stimulus presentation on transposition. *J. comp. physiol. Psychol.*, 1951, **44**, 378–382.
2. Grice, G. R. Visual discrimination learning with simultaneous and successive presentation of stimuli. *J. comp. physiol. Psychol.*, 1949, **42**, 365–373.
3. Hunter, I. M. L. The absolute and relative theories of transposition behavior in rats. *J. comp. physiol. Psychol.*, 1953, **46**, 493–497.
4. Köhler, W. *Gestalt psychology*. New York: Liveright, 1929.
5. Lashley, K. S. The mechanism of vision. XV. Preliminary studies of the rat's capacity for detail vision. *J. gen. Psychol.*, 1938, **18**, 123–193.
6. Lashley, K. S., & Wade, M. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.
7. Nissen, H. W., Levinson, B., & Nichols, J. W. Reinforcement and "hypothesis" in the discrimination behavior of chimpanzees. *J. exp. Psychol.*, 1953, **45**, 334–340.
8. Saldanha, E. L., & Bitterman, M. E. Relational learning in the rat. *Amer. J. Psychol.*, 1951, **64**, 37–53.
9. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.
10. Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.
11. Webb, W. B. A test of "relational" vs. "specific stimulus" learning in discrimination problems. *J. comp. physiol. Psychol.*, 1950, **43**, 70–72.
12. Weiss, E. S., & Stevenson, H. W. Time as a variable in transposition. *Amer. J. Psychol.*, 1955, in press.

(Received October 8, 1954)

Journal of Experimental Psychology
Vol. 50, No. 3, 1955

MOVEMENT ERROR, PRESSURE VARIATION, AND THE RANGE EFFECT[1]

BERNARD WEISS[2]

University of Rochester

A motor response that cannot be guided by external (visual) feedback may be guided by internal (proprioceptive) feedback. In an earlier study (**8**), proprioceptive feedback was systematically varied by varying the maximum pressure and displacement needed to perform a stick positioning task. The task required adjustments to be made after the original visual cues were withdrawn. Two sets of pressure-displacement conditions were used to obtain the variation in pressure and displacement. In one set of conditions the maximum pressure required on the control stick was kept constant at 30 lb. and the maximum displacement was varied, taking on values of 3, 6, 12, 18, 24, and 30 degrees. In the other set, maximum displacement was kept constant at 30° and maximum pressure was varied from 0 to 30 lb., in 6-lb. steps. Pressure variation produced no significant effect on constant error (CE) and variability in the set of conditions in which displacement remained constant and pressure was varied. When, on the other hand, pressure remained constant and displacement was varied, CE and relative variability decreased as the maximum displacement increased. The experimental design, however, did not permit a definite function to be assigned to pressure cues in the positioning settings made with the smaller displacement ranges (less than 7.5°) since maximum pressure for these ranges was constant.

The earlier study also showed that the differences among these conditions mainly arose from responses to the shortest displacement within a condition. This result was attributed partly to the "range effect," a phenomenon observed by many psychologists, and investigated in motor performance by Ellson and Wheeler (**2**), Ellson and Coppock (**3**), and Slack (**6**), among others. The range effect describes a tendency for *S*s to overestimate a particular displacement when it is the shortest in a range of stimuli, and to underestimate it when it is the longest.

The present study, therefore, had two purposes: first, a more precise investigation of the role played by pressure variation in positioning settings with displacement ranges less then 7.5°; second, some assessment of the extent to which the range effect had produced the results of the previous study (**8**).

METHOD

Four conditions were used. In each condition the maximum displacement required was 6°. The different conditions were distinguished by their maximum pressures. These four values were 0, 7.5, 15, and 30 lb. Pressure and distance were linearly related.

The *S*'s task, as in the previous study, was to compensate for the displacement of a spot of light from the center of an oscilloscope screen. The spot flashed on for .6 sec., then was withdrawn for 2.5 sec. During the "blackout" period, *S* moved a 20-in. stick control the distance and in the direction (either fore or aft)

[1] This study was conducted as part of a research program on human motor performance supported by a contract between the Special Devices Center, Office of Naval Research, and the University of Rochester, with Dr. S. D. S. Spragg as responsible investigator.

[2] Now at the USAF School of Aviation Medicine, Randolph Field, Texas.

TABLE 1

CONSTANT ERROR IN DEGREES FOR THE MIXED AND CONSTANT SERIES AS A FUNCTION OF ANGLE AND PRESSURE

Angle	Constant Series					Mixed Series					Mean (Both Series)
	0 Lb.	0–7.5 Lb.	0–15 Lb.	0–30 Lb.	Mean	0 Lb.	0–7.5 Lb.	0–15 Lb.	0–30 Lb.	Mean	
1.5°	.038	.122	.211	−.015	.089	.241	.377	.307	.413	.334	.212
3°	.153	.158	.094	.001	.102	.011	.040	.024	.130	.051	.076
4.5°	.170	.085	.063	−.110	.052	−.132	−.341	−.211	−.135	−.205	−.076
6°	−.077	−.025	.088	−.030	−.011	−.071	−.142	−.342	−.450	−.251	−.131
Mean	.071	.085	.114	−.038	.058	.012	−.016	−.056	−.011	−.018	.020

that he expected would return the spot to the center of the screen. At the end of this period, the spot reappeared for 2 sec. to give *S* an opportunity to correct his error. After a 5-sec. "blackout" interval, another displacement appeared on the screen. A more detailed description of the apparatus and task is given elsewhere (**8**, p. 216).

There were four possible displacements in each direction (push and pull). These corresponded to 1.5, 3, 4.5 and 6 degrees, or 25, 50, 75, and 100% of the maximum distance.

Subjects.—All 8 *S*s were graduate students or staff members in the psychology department of the University of Rochester and ranged in age from 23 to 32 years. None of them had served as *S*s in the previous study.

Procedure.—To obtain a measure of the influence of the range effect, two series of displacements were used. In the mixed series, the sequence of presented displacements was essentially random, as was that used in the previous study (**8**). In the constant series, all the displacements in a sequence were the same size.

Four sessions were devoted to the mixed and four to the constant series for each *S*. Thus, each *S* performed on each pressure-displacement condition under two different methods, making a total of eight experimental sessions per *S*.

On the mixed series, a trial consisted of 32 displacements. Since there were 8 possible displacements (4 in each direction), each trial consisted of 4 settings for each displacement. Each session was made up of 4 trials, so that during a single session *S* made 16 settings for each of the 8 possible displacements.

Since it could be expected that practice on a series of displacements of constant size would tend to eliminate the range effect (**3**), it was possible, by comparing performance on the mixed with performance on the constant series, to obtain an estimate of how much the range effect had contributed to the earlier results.

To keep the amount of practice on a particular size of displacement for a particular condition equivalent for both series, the length of each sequence in the constant series was chosen to be 16 settings. Order effects were counterbalanced by a latin-square design.

As in the previous experiment, the *S*s received preliminary training to acquaint them with the apparatus and task. At the beginning of each session, *S* made one setting at each of the 8 displacements, which were flashed on manually by *E*. For the rest of the session, the displacements were presented automatically by a selective sequence programmer and timer. The presented displacement and *S*'s response were recorded by a Brush BL-202 inkwriter oscillograph.

Each *S* performed on all conditions with both series. The effects of sequence of conditions were controlled by counterbalancing.

RESULTS

For the mixed series, the results to be analyzed are derived from the last trial of each session. Since each trial

TABLE 2

ANALYSIS OF VARIANCE OF THE CONSTANT ERROR DATA

Source	*df*	Mean Square	*F*
Pressure	3	7,356.78	.53
Angle	3	192,119.87	13.71*
Series	1	45,375.78	3.26
P × A	9	6,730.70	.48
P × S	3	13,738.95	.99
A × S	3	109,132.03	7.85*
P × A × S	9	14,011.75	
Total	31		

* $P < .01$.

TABLE 3

WEBER RATIO FOR THE MIXED AND CONSTANT SERIES AS A FUNCTION OF ANGLE AND PRESSURE

Angle	Constant Series					Mixed Series					Mean (Both Series)
	0 Lb.	0–7.5 Lb.	0–15 Lb.	0–30 Lb.	Mean	0 Lb.	0–7.5 Lb.	0–15 Lb.	0–30 Lb.	Mean	
1.5°	.158	.207	.226	.139	.182	.267	.351	.349	.352	.330	.256
3°	.183	.132	.097	.083	.124	.171	.195	.198	.172	.184	.154
4.5°	.124	.123	.078	.093	.104	.188	.172	.185	.147	.173	.138
6°	.101	.087	.102	.091	.095	.129	.159	.135	.142	.141	.118
Mean	.142	.137	.126	.102	.126	.189	.219	.217	.203	.207	.166

contained 4 settings for each of the 8 displacements, every displacement within a condition was defined by data from 32 settings or 4 by each *S*.

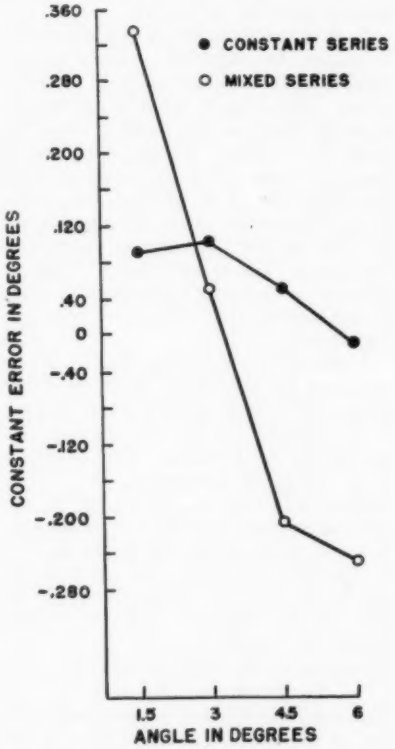

FIG. 1. Constant error in degrees for the mixed and constant series as a function of angle.

The last 4 settings of each sequence in the constant series were selected to be compared to the comparable settings on the last trial of the mixed series, since the latter comprised the last 4 of the 16 settings that were made during a session on each of the 8 displacements.

The results are described in two ways: constant error and Weber ratio, the latter being computed as the ratio of the *SD* to the size of the required displacement. The data for push and pull are combined since the differences between the directions were not significant.

Constant errors.—The CE's in degrees for both series are given in Table 1. The significance of the observed differences was assessed by

TABLE 4

ANALYSIS OF VARIANCE OF THE WEBER FRACTION DATA

Source	*df*	Mean Square	*F*
Pressure	3	964.75	1.62
Angles	3	30,106.42	50.61*
Series	1	51,842.00	87.15*
P × A	9	945.50	1.59
P × S	3	1,113.08	1.87
A × S	3	4,133.25	6.95**
P × A × S	9	594.89	
Total	31		

* $P < .01$.
** $P < .02$.

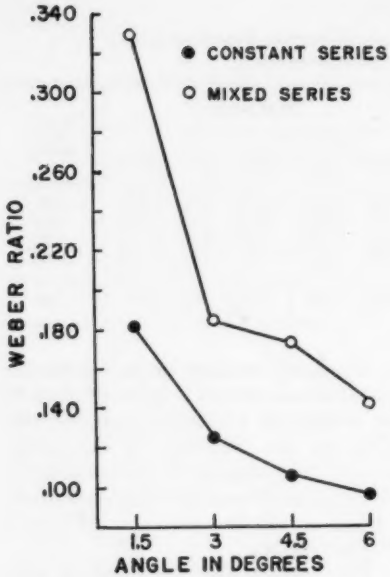

FIG. 2. Weber ratios for the mixed and constant series as a function of angle.

an analysis of variance, shown in Table 2. This reveals that the differences among pressures and series are not significant by the usual criteria. The differences among angles are significant. As can be seen in Fig. 1, the significance is mainly derived from differences on the mixed series, which shows a prominent "range effect." This statement is supported by the significant Angle × Series interaction, which also means that the slopes of the two curves are significantly different.

Weber ratios.—Table 3 gives the Weber ratios for the two series and Table 4 shows the results of the analysis of variance. Again, pressure differences are not significant, although a slight trend appears in the data for the constant series. Angles and Series, however, did produce significant variation. Figure 2 shows that except for the initial rate of decline from 1.5 to 3 degrees, the Weber ratios for the two series are proportional. The F ratio for the Angle × Series interaction, however, indicates that these initial slopes are significantly different.

DISCUSSION

Three conclusions seem to follow from the results of this and the earlier study (8):

1. When a positioning task is used in which visual feedback is delayed, pressure variation, over a wide range of pressure and displacement values, will not significantly alter CE or the precision of adjustment as defined by the Weber ratio or relative variability.

2. A mixed sequence of displacements, as might be expected, results in less precise discrimination than a constant series. But, also, it tends to produce its greatest effect at the shortest displacement in a range, making the loss of precision here proportionately greater (see Fig. 2). The relative variability data of the earlier study (8) also suggest that the shorter the range of movement, the greater the initial slope. That is, shortening the range of movement tends to exaggerate the departure from the horizontal.

3. Similarly, there seems to be a tendency for the range effect, as it appears in this task (on the mixed series), to be skewed toward the shorter displacements in a range. That is, for a single range of movement, the CE in degrees will tend to be higher at the shortest displacement than at the longest. A rough index of this skewness may be secured by computing the ratio of the CE at the shortest displacement to the CE at the longest. For the earlier study, the mean ratio for the six conditions in which displacement varied is 1.37. In the present study it is 1.33. Also, the ratio appears to change with the size of the range, growing larger as the range becomes smaller, and thus dis-

playing a trend similar to that found for precision.

Possible explanations for the apparent ineffectiveness of pressure variation as a cue in this kind of task were discussed in the earlier paper (**8**) and seem applicable here. Two alternatives were offered. One was that distance discrimination may be so much finer than pressure discrimination that most of the pressure information is redundant. The other was that distance information may be remembered more easily and more precisely than pressure information, since it seems easier to conceptualize and provides a greater variety of cues.

The finding that CE on a constant series of misalignments is less than on a mixed series is hardly an occasion for surprise. Other investigators have reported similar results when using tasks like that used here (**1, 3**). The interest it holds arises from the difference in the pattern of error, which was reflected in the significant Angle × Series interaction and in the ratio of CE at the shortest to CE at the longest displacement. Perhaps the phenomenon taking place might be called a "nonlinear range effect." Other data also suggest that with mixed series such an effect may be obtained. Slack's calculations (**6**), for instance, reveal that the responses to the smallest displacement in his task are more heavily skewed toward the mean of the input distribution than responses to the larger ones.

Models to account for the range effect, which Hollingworth called the law of central tendency (**4**), have been proposed by Johnson (**5**) and by Slack (**7**). Johnson views this phenomenon as another manifestation of the regression toward the mean that occurs when two variables are imperfectly correlated—the variables in this instance consisting of stimulus magnitudes and response magnitudes. Slack's model proposes that the error on any trial will be a function of a weighted average of the differences among succeeding magnitudes of the earlier stimuli. Both of these models assume linearity of the effect.

The Weber fraction values provide an interesting corollary to the CE data. They show, not only that a mixed series produces greater dispersion than a constant series, but that, as with CE, the pattern, or distribution of variable error, i.e., Weber ratio, is altered. Perhaps more important is the fact that for both series the Weber fraction shows a progressive decline with angle. That this is not a phenomenon restricted to one particular range of movement may be seen by examining the relative variability data of the earlier experiment (**8**), which indicate that the effect is more prominent with the shorter ranges of movement. It appears, therefore, to be a phenomenon different from the decrease in precision at low-intensity values often associated with the Weber functions of various sensory modalities. Perhaps it is not too unreasonable, in fact, to suspect that part of the latter findings may be due to this "skewed dispersion effect." In any event, it provides one more complication for models of psychophysical judgment to account for.

Summary

The *S*'s task in this study was to compensate for the displacement of a spot of light from the center of an oscilloscope screen. He did this by moving a control stick during a period when the spot was not visible. To determine the effect of pressure variation on the CE and precision (Weber ratio) of positioning responses at small distances, four pressure-displacement conditions were used: a maximum displacement of 6° with maximum pressures of 0, 7.5, 15, and 30 lb. Eight *S*s were used, each serving under all conditions. The displacements were presented in two orders: a mixed (random) series and a constant series.

For both series, pressure variation had no significant effect on CE and the Weber fraction. The pattern of error and dispersion for the two series differed, however. In the mixed series, the range effect was quite prominent as compared to the constant series. Also, the CE at the shortest displacements, when the data of an earlier experiment are included, seems to be greater than the CE at the longest, thus indicating nonlinearity of the range effect. This nonlinearity seems to be exaggerated as the range of movement is shortened. The Weber

ratios for the two series seem to be proportional except for the initial decline as the extent of movement increases, the mixed series resulting in a steeper fall. Both series show a greater Weber ratio at the shorter displacements, and along with the data of the earlier study suggest that this departure from linearity is magnified by compressing the range of movement.

References

1. Craik, K. J. W., & Vince, M. Psychological and physiological aspects of control mechanisms, Part II, with special reference to ground-tank and A.A. tank gunnery. Medical Research Council, Military Personnel Research Committee, Great Britian, BPC 44/322, 1944.
2. Ellson, D. G., & Wheeler, L. The range effect. Dayton, O.: Wright-Patterson Air Force Base, 1951. (*Tech. Rep.* 5813).
3. Ellson, D. G., & Coppock, H. Further analysis of the psychological range effect. Dayton, O.: Wright-Patterson Air Force Base, 1951. (*Tech. Rep.* 6012).
4. Hollingworth, H. L. The inaccuracy of movement with special reference to constant errors. *Arch. Psychol., N.Y.*, 1909, **2**, No. 13.
5. Johnson, D. M. The central tendency of judgment as a regression phenomenon. *Amer. Psychologist*, 1952, **7**, 281. (Abstract)
6. Slack, C. W. Some characteristics of the "range effect." *J. exp. Psychol.*, 1953, **46**, 76–80.
7. Slack, C. W. Perception and past experience: a mathematical model for Taylor's "range effect." *Amer. Psychologist*, 1953, **8**, 437. (Abstract)
8. Weiss, B. The role of proprioceptive feedback in positioning responses. *J. exp. Psychol.*, 1954, **47**, 215–224.

(Received October 25, 1954)

Journal of Experimental Psychology
Vol. 50, No. 3, 1955

GENERALIZED EXTINCTION AND SECONDARY REINFORCEMENT IN VISUAL DISCRIMINATION LEARNING WITH DELAYED REWARD

G. ROBERT GRICE AND HERBERT M. GOLDMAN

University of Illinois

In a study of the role of secondary reinforcement in delayed reward learning, Grice (1) has found that white-black discrimination learning is facilitated by the use of goal boxes of the same color as the alleys comprising the choice stimuli. In an experimental group, animals entered either a black or white stimulus alley, and, after a 5-sec. delay in a gray compartment, entered an end box of the same color as the stimulus chosen. The reward was received in the white end box. For the control group, both end boxes were gray, the reward being placed in the gray box following choices of the white alley. The superiority of the experimental group was attributed to the generalization of secondary reinforcement from the white goal box to the white stimulus alley. This was assumed to provide a basis for immediate reward following the correct choice which was not present for the control group. This interpretation neglects, however, the possible effects on learning of the fact that incorrect choices always led to an empty end box of the same color as the negative stimulus alley. It is quite possible that inhibitory effects of the empty end box may generalize and provide a basis for rapid extinction of approach responses to the negative stimulus alley. This possibility is readily deduced from S–R interpretations of discrimination learning which emphasize the role of inhibitory factors (**4, 6**), and is also suggested by experimental findings indicating their importance (**2, 3, 5**).

The purpose of the present experiment is to evaluate the relative contributions of generalized secondary reinforcement and inhibition at the choice point resulting from the manipulation of end-box experiences.

PROCEDURE

General design.—The experiment involved four groups trained on a white-black discrimination problem with 2-sec. delay following choice of either the correct or incorrect alley. The independent variable was the color of the end boxes following either choice. As in the earlier study (1), all *S*'s were trained with white positive. For the control group (GG) both of the end boxes were gray, the food being placed in the box following the white stimulus alley. For a second group (WB), there was a white goal box and a black, negative end box. This condition corresponds to that of the original experiment and superiority over the control condition may be attributed either to generalized secondary reinforcement or inhibition or both. A third group (WG) had a white goal box and a gray, negative end box. This condition presumably involves the secondary reinforcement effect but not the inhibition effect. The fourth group (GB) had a gray goal box and a black, negative end box. Superiority of this group over the control should be due to the generalization of extinction but should not involve the secondary reinforcement effect.

Apparatus.—The apparatus was the black and white alley, delayed reward discrimination apparatus previously described by Grice (1). There were certain minor differences in the present form of the apparatus: (*a*) The floor of the V-shaped choice chamber was gray rather than extensions of the black and white alleys. (*b*) The color of the end compartments was varied by the use of sliding three-alley sections similar to the stimulus alleys rather than by the use of inserts. (*c*) Lighting was from bulbs placed over the stimulus alleys and over the end boxes. The average brightnesses of the white stimulus alley and the white goal boxes were

TABLE 1

TRIALS AND ERRORS TO CRITERION OF LEARNING

Group	Errors			Trials		
	Median	Mean	Range	Median	Mean	Range
GG	128.5	*	100–492*	305	*	240–880*
WB	43	38.9	12–59	105	103	30–190
WG	58	61.5	27–93	165	158	80–230
GB	40	40.6	11–79	110	106	30–220

* Means are indeterminate since one *S* in the GG group failed to learn in 880 trials and was discontinued.

approximately 25.6 and 23.6 apparent ft.-candles, respectively. The brightness of the black stimulus alley was about 1.3 apparent ft.-candles, and that of the black end box was 1.1. The gray goal boxes were about 5.0 apparent ft.-candles. Each of these values is a mean of 10 readings with a Macbeth illuminometer. (*d*) The 2-sec. delay employed with all groups in the present experiment was achieved by delaying *S*s for 2 sec. in the 18-in. delay compartments rather than by allowing them to run through 6-ft. alleys as in the previous experiment. Timing was done with a stop watch.

Subjects.—The *S*'s were 20 male and 20 female, experimentally naive, albino rats of Sprague-Dawley strain. Their ages ranged from 73 to 98 days at the beginning of the experiment. The *S*'s were randomly assigned to the groups with the restrictions that each group contain the same number of males and females and that litter mates be divided equally among the groups whenever possible. Three animals originally assigned to groups were discarded and replaced. Two were eliminated because of experimenter error and the other because it displayed an initial preference for the positive stimulus (9 out of the first 10 trials).

Training procedure.—The pretaining and training procedures have been previously described (1). The only departure from the previous procedure was that in the black-white discrimination training, there were 10 trials per day for only the first 40 trials instead of 200, and, after that, there were 20 trials per day. The criterion of learning was 18 out of 20 correct responses with the last 10 perfect in two successive blocks of 10 trials.

RESULTS AND DISCUSSION

The mean and median number of trials to criterion and number of total errors are presented in Table 1. The Mann-Whitney *U* test of significance was applied to the differences between the groups and the results of these tests are presented in Table 2. The group learning curves based on percentage correct for blocks of 20 trials are presented in Fig. 1. In order to plot these curves, *S*s were assumed to run 100% correct after the criterion runs.

The most obvious finding is that all three of the groups involving differential end-box stimulation are clearly superior to the GG control group. There was no overlap in either trials or errors between this group and any of the three experimental groups. Differences among the three experimental groups are small. The White-Black and Gray-Black groups are approximately equal and both are somewhat superior to the White-Gray condition.

The data clearly indicate the effectiveness of the black end box following the

TABLE 2

RESULTS OF MANN–WHITNEY *U* TEST OF DIFFERENCES BETWEEN GROUPS

Groups Compared	Errors *P*	Trials *P*
GG vs. WB	<.0001	<.0001
GG vs. WG	<.0001	<.0001
GG vs. GB	<.0001	<.0001
WB vs. GB	.409	.224
WB vs. WG	.035	.019
WG vs. GB	.045	.145

choice of the negative, black alley. In fact, this condition produced just as rapid learning as did the presence of end boxes resembling both of the stimulus alleys. This finding indicates that Grice's analysis of the original experiment (1) was incomplete, and suggests that an important factor in such learning may be the generalization of extinction from the negative end box to the similar alley at the choice point. This would then tend to produce a more rapid elimination of errors than in the control group. To extend the original theoretical analysis of delayed reward learning, this finding suggests that learning with delayed reward will be facilitated not only by conditions producing immediate secondary reinforcement of the correct response, but also by conditions hastening extinction of competing responses.

While apparently incomplete, Grice's original analysis in terms of secondary reinforcement is further supported by the present experiment. The clear-cut superiority of the White-Gray group over the control group indicates that the generalization of secondary reinforcement to the positive choice-point stimulus does facilitate learning even in the absence of the generalized inhibitory effect of the black end box. The data, then, are consistent with the original hypothesis of Spence (7) and Grice (1) that delayed reward learning is produced by immediate secondary reinforcement.

In view of the fact that generalized secondary reinforcement and generalized extinction are both effective in the present situation (as evidenced by the WG and GB groups), it may appear rather strange that these two effects did not clearly summate to produce a marked superiority of the WB group where presumably both are involved. The WB group was somewhat superior to the WG group but about the same as the GB group. Of course, the logic of statistical reasoning does not allow us to conclude that no true superiority of the WB condition exists, but even so, this failure to obtain a difference is not as strange as it first appears. First, we may assume that the habits of approaching both the black and white alleys attain considerable strength as the result

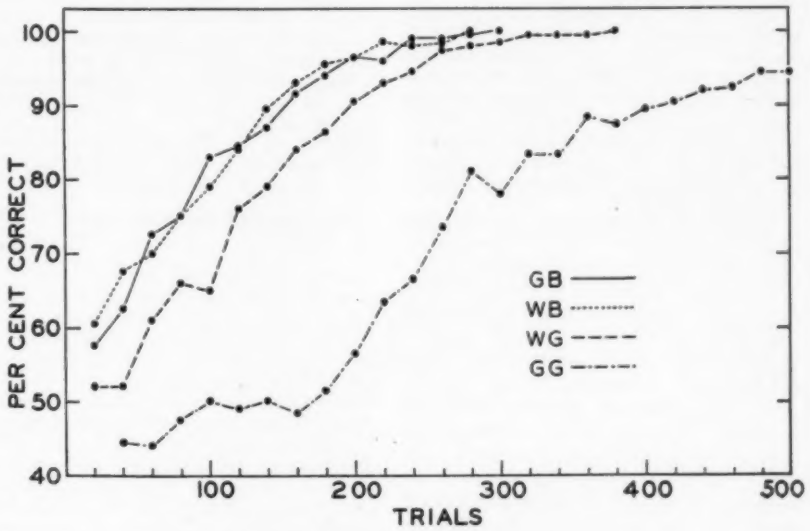

FIG. 1. Group learning curves in 20-trial blocks.

of the *nondifferential* secondary reinforcement from entry into the gray delay compartments. Thus, the added reinforcement received from the use of the white goal box would add rather small increments to the habit under the assumption of a negatively accelerated habit growth curve. It is not unreasonable, then, that the generalized inhibitory effect upon approach responses to the black should produce a more rapid elimination of errors. This is indicated by the superiority of the GB group over the WG. Since the extinction effect occurs in fewer trials than the reinforcement effect, we are led to believe that the WB and GB groups learned essentially in the same way, namely, by the extinction of responses to the negative stimulus. A more complete analysis of the situation would indicate that, even in view of these considerations, there should be some superiority of the combined effect in the WB group. However, its superiority would be small and perhaps clearly demonstrable only in a very large-scale experiment.

Summary

Four groups of rats were run in a white-black visual discrimination learning situation with 2-sec. delay of reward. In one group, the end boxes were the same in color as the corresponding stimulus alleys. For a second group only the goal box was the same as the positive choice-point stimulus. In the third group, the empty end box was the same as the negative stimulus. For a control group neither end box was like the choice-point stimuli.

In terms of trials and errors to a criterion, all three experimental groups were significantly superior to the control group. There were only small differences among the three experimental groups, but the group with only the goal box similar to the positive stimulus was somewhat inferior to the other two.

The results substantiate an earlier hypothesis explaining delayed reward discrimination learning in terms of generalization of secondary reinforcement to the choice-point cue. However, this interpretation was found to be incomplete, and the experiment suggests that in this situation an even more effective variable is the generalization of extinction to the negative choice-point cue.

References

1. Grice, G. R. The relation of secondary reinforcement to delayed reward in visual discrimination learning. *J. exp. Psychol.*, 1948, **38**, 1–16.
2. Grice, G. R. Visual discrimination learning with simultaneous and successive presentation of stimuli. *J. comp. physiol. Psychol.*, 1949, **42**, 365–373.
3. Grice, G. R. The acquisition of a visual discrimination habit following extinction of response to one stimulus. *J. comp. physiol. Psychol.*, 1951, **44**, 149–153.
4. Hull, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1939, **46**, 9–30.
5. Kling, J. W. Generalization of extinction of an instrumental response to stimuli varying in the size dimension. *J. exp. Psychol.*, 1952, **44**, 339–346.
6. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.
7. Spence, K. W. The role of secondary reinforcement in delayed reward learning. *Psychol. Rev.*, 1947, **54**, 1–8.

(Received November 1, 1954)

Journal of Experimental Psychology
Vol. 50, No. 3, 1955

THE PREDICTION OF DECISIONS AMONG BETS [1]

WARD EDWARDS

Armament Systems Personnel Research Laboratory, Air Force Personnel and Training Research Center, Lowry AFB, Denver

This paper presents a very simple mathematical model for predicting choices among bets and an experiment in which that model is tested. The model is based on the concepts of utility and subjective probability and on the theory of games.

Development of the Model

Four models for decision-making.—This paper will discuss four similar simple models which have been used to make predictions of choices among bets. They are defined by the following four equations:

$$EV = \sum_i p_i \$_i \qquad (1)$$

$$EU = \sum_i p_i u_i \qquad (2)$$

$$SEM = \sum_i P^*_i \$_i \qquad (3)$$

$$SEU = \sum_i P^*_i u_i \qquad (4)$$

Equation 1 defines the concept of *expected value* (*EV*). The symbol p_i stands for the probability of the ith outcome of the bet, the symbol $\$_i$ stands for the value of the ith outcome of the bet in dollars, and $\sum_i p_i = 1$. The *EV* of a bet is the amount which a gambler will receive as a result of making it, on the average. Consequently, the traditional normative theory about choices among bets is that gamblers should choose the bet with the highest *EV* or lowest negative *EV*. (A normative theory is a theory about what people should do, rather than what they do do.)

It is, of course, apparent that people don't do this, and it is even doubtful in some cases that they should. Consequently, Bernoulli (1) suggested instead that gamblers should choose the bet with the highest *expected utility* (*EU*), where *EU* is defined as in Equation 2. The symbol u_i in Equation 2 is the *utility*, or subjective value, of the ith outcome of the bet. This proposal, then, simply is a suggestion that the utility or subjective value of a given amount of money be substituted for its objective value in the calculation of an expected value.

Recently, von Neumann and Morgenstern (13) have revived Bernoulli's suggestion, and have used Equation 2 as a definition of utility. This assumes, of course, that *EU* maximization is a theory about what people do do, rather than a normative theory. Following von Neumann and Morgenstern's suggestion, Mosteller and Nogee (10) assumed the validity of the *EU* maximization model and used it to measure the utility of small amounts of money experimentally.

If it is reasonable to assume that subjective values of money should be substituted for objective values in Equation 1, it is equally reasonable to make the same assumption about probabilities. The notion of subjective probabilities has been discussed for some time; one of the most explicit uses of it (without use of the concept of utility) was made by Preston and Baratta (11). They implicitly suggested that *S*s be assumed to make choices among bets so as to maximize what might be called *sub-*

[1] This research was done while the author was at The Johns Hopkins University. It was done under Contract N5-ori-166, Task Order 1, between the Office of Naval Research and The Johns Hopkins University. This is Report No. 166-I-191, Project Designation No. NR-145-089, under that contract. I am indebted to Miss Frances Brown, who ran *S*s and processed much of the data, and to Miss Susanne Wollenberg, who also helped with the data processing. A very preliminary report of this experiment, including only the data from *S*s 3 and 5, was presented to the Dunlap Symposium on Mathematical Models in the Social Sciences (8).

jectively expected money value (*SEM*), which is defined by Equation 3. The symbol P^*_i in Equation 3 stands for the subjective probability corresponding to the objective probability of the ith alternative. No addition theorem applies to the P^* values; however, the sum of the objective probabilities corresponding to the various P^* values in Equation 3 is 1. Having assumed the validity of the *SEM* maximization model, Preston and Baratta used it to calculate a subjective probability scale. Mosteller and Nogee also used Equation 3 to calculate subjective probabilities, but preferred the utility interpretation given by Equation 2. Similarly, it would be possible to analyze Preston and Baratta's data by Equation 2 instead of Equation 3. There is little to choose between these opposed interpretations of either of these two experiments.

In a series of experiments Edwards (**3, 4, 5**) has shown that *S*s prefer to bet at some probabilities rather than others, that these preferences are essentially independent of the amounts of money involved so long as they are not too large and so long as all bets compared with each other have the same *EV*, and that *S*s prefer less favorable bets at their preferred probabilities to more favorable bets at disliked probabilities. These experiments show that there are two patterns of probability preferences, one for positive *EV* bets and another for negative *EV* bets. These results, of course, can be easily accounted for by means of the *SEM* model, but are inconsistent with the *EU* model.

All these considerations suggest a more comprehensive model, which includes both utility and subjective probability. This paper is intended to present and test such a model. The model asserts that people choose among bets so as to maximize *subjectively expected utility* (*SEU*), which is defined in Equation 4. The symbols in Equation 4 have the same meaning as before. This model, like the others, does not assert that *S*s know the *SEU* values of various bets and deliberately choose the bet with the largest *SEU*; it asserts only that the hypothesis of *SEU* maximization will enable an *E* to predict the choices that *S*s will actually make.

This model is similar to one proposed by Coombs and Beardslee (**2**). The most important difference is that Coombs and Beardslee assume both P^* and u to be measurable only on an ordered metric, while this model (and the *EU* and *SEM* models also) assume them to be measurable on an interval scale.

For a fuller discussion of the concepts of utility and subjective probability and of the Mosteller and Nogee, Preston and Baratta, and other related experiments, see (**6**).

The Test of the Model

In order to test the *SEU* maximization model, three things are necessary. Scales of utility and of subjective probability must be determined for *S*s whose behavior is to be predicted, and *S*s must make choices among bets, so that these choices can be compared with those predicted by the model.

The crux of the problem of testing this model is to devise methods of measuring utility and subjective probability. Classical psychophysical methods seem inappropriate; both probabilities and amounts of money come stated in numerical form, and psychophysical judgments about them are almost certain to be made in terms of their numerical rather than their subjective values.

The matrix method.—The theoretically ideal method of measuring utility and subjective probability simultaneously in a gambling situation has been suggested by Woodbury (**14**). The essence of it is that a rectangular matrix with amounts of money on one margin and probabilities on the other defines a series of probability-amount combinations, or bets. An *S* can be required to inform *E* of amounts of money such that he is indifferent between them and the bets in the matrix. The *SEU* maximization model permits a system of simultaneous equations based on these judgments to be set up, enough

more equations can be added by linear interpolation between adjacent utility values to make the system soluble, and the solution gives the utility and subjective probability functions needed.

This procedure was tried in this experiment, using the method of adjustment. The data were so variable as to be meaningless, and so will not be further discussed. Less variable data could probably be gotten by the method of constant stimuli, but it is unlikely that they would be good enough to meet the exacting needs of the matrix method.

The N-bets method of utility measurement.—If no adequate technique for measuring utility and subjective probability simultaneously can be devised, then it is necessary to measure them separately. The methods used by Mosteller and Nogee (10) and by Preston and Baratta (11) are inappropriate, since they irretrievably confound utility and subjective probability. Some technique must be invented which either eliminates one of these variables or else holds it constant.

The method used in this experiment, which will hereafter be called the *N*-bets method, is based on a sweeping, convenient assumption. That assumption is that any bet can be considered as an object which has a utility (which can be calculated from Equation 4, once the utility and subjective probability functions are known), and that the utility of N identical bets is N times the utility of one such bet. This is an assumption that what the economists call the *marginal utility* function for bets, but not necessarily for any other object, is a constant. (The marginal utility of a commodity is the first derivative of its total utility curve.)

How can this assumption be used to measure utility? Suppose that S is indifferent between one 50–50 bet in which he wins \$A if he wins and nothing if he loses, and N 50–50 bets in each of which he wins \$0.10 if he wins and nothing if he loses. The assumption discussed above makes it legitimate to write:

$$P^*_{0.5}\, u_{\$A} + P^*_{0.5}\, u_{\$0.00} = N\,(P^*_{0.5}\, u_{\$0.10} + P^*_{0.5}\, u_{\$0.00}) \quad (5)$$

Since both utility and subjective probability are defined as measurable on an interval scale, we may arbitrarily assign two points on each curve. Here and throughout this paper the following arbitrary definitions will be used: $u_{\$0.00} = 0$; $u_{+\$0.10} = 10$; $u_{-\$0.10} = -10$; $P^*_0 = 0$; $P^*_1 = 1$. In cases where \$0.15 small bets are used, $u_{+\$0.15} = 15$ and $u_{-\$0.15} = -15$.

If these definitions and some simple algebra are applied to Equation 5, the terms involving $u_{\$0.00}$ drop out, the $P^*_{0.5}$ factors out, and Equation 5 simplifies to:

$$u_{\$A} = 10\,N \quad (6)$$

Since N is known, the utility of \$A can be calculated. The same procedure can be carried out for other amounts of money, and so the whole utility curve can be determined. It should be noted that this utility curve, like those determined by Mosteller and Nogee, is not a total utility curve of the type which is usual in economics. Its zero point is where S is at the moment he makes his choice. Since S wins and loses money during the course of each session, that zero point is not a fixed point on his total utility curve in the classical sense. Consequently, the kind of utility curve which this procedure produces is closely analogous to the kind suggested by Markowitz (9), except that Markowitz takes zero to be S's usual location, while this procedure takes it to be his current location.

The *N*-bets method offers two advantages. First, it is straightforward and the experiment it suggests is easy to perform. Secondly, it offers a built-in check on some of the assumptions involved. There is no reason why 50–50 bets need be used, and there is no reason why the small bets should involve \$0.10. Other probabilities and other sizes of small bets could be used. All probabilities and all sizes of small bets should lead to identical utility curves. If they do not, then it probably means that the *N*-bets assumption is incorrect, and the marginal utility of bets is not constant. For any given marginal utility function for bets, it is easy to predict the exact nature of the difference between utility

functions obtained from different sizes of little bets.

If *S*s preferred to bet at some variances rather than others, then the existence of such preferences would invalidate the *N*-bets method of utility measurement, and also the *EU*, *SEM*, and *SEU* maximization models. However, Edwards (7) has shown that such preferences, though they do exist, are of minute importance as compared with probability preferences, and so can be neglected.

How to measure subjective probabilities.—Once a utility scale is available for a given *S*, it is very easy to measure subjective probabilities. Suppose that *S* is indifferent between a certain gain of $0.55 and a 3/8 chance of winning $2.50. The *SEU* maximization model permits us to write:

$$P^*_{3/8}\, u_{\$2.50} + P^*_{5/8}\, u_{\$0.00} = P^*_1\, u_{\$0.55} \quad (7)$$

This simplifies to:

$$P^*_{3/8} = u_{\$0.55}/u_{\$2.50} \quad (8)$$

Since the two utilities are known, the value of the subjective probability corresponding to an objective probability of 3/8 can be calculated. By determining other indifference points involving other probabilities, other subjective probabilities can be calculated. No assumptions beyond those involved in the *SEU* maximization model are necessary. The same procedure, carried out with losses instead of gains, gives the subjective probability scale for losing bets.

How to test the SEU maximization model.—Once individual utility scales and individual subjective probability scales are available, testing the *SEU* maximization model is accomplished by having the same *S*s make choices among bets. The *SEU* of each bet for each *S* can be calculated, and the prediction is that he will always choose the bet with the higher *SEU* (or smaller negative *SEU*, if no positive *SEU* is available). Since utility and subjective probability are measured only on interval scales, they have no true zero points. Consequently, the *SEU* model cannot be used to predict choices among bets which include both possibilities of winning and possibilities of losing. So simple bets, in which only possibilities of winning or possibilities of losing appear, must be used to test the *SEU* model.

Method

Apparatus.—All gambling in this experiment was done on two pieces of gambling apparatus. One was a pinball machine, which has been more fully described elsewhere (3). It has eight cells along its bottom, into any one of which the ball may roll. The *S*s are told that the ball is just as likely to roll into any one cell as into any other. Under the surface of the machine eighteen electromagnets are concealed, which completely control which cell the ball ends up in. These electromagnets are operated by a programmer in the next room, and the program is set up so that the instructions to *S*s are correct.

The other apparatus was designed because it seemed desirable to have some apparatus which could run off the strings of identical bets necessary for the *N*-bets method of utility measurement considerably more quickly than they could be run off on the pinball machine. Consequently, a bank of eight counters was set up. The counters were numbered from 1 through 8, and were automatically resettable. In front of them was a button. When *S* pushed it, one of the eight counters received one electrical impulse. A programmer determined which counter would be operated. By holding the button down, *S* could operate the apparatus continuously, so that a succession of impulses operated various counters. An auxiliary counter counted the total number of impulses entered into the bank of eight counters since the last resetting. Thus, in order to run off 20 identical bets, all *S* had to do was to hold down the button until the auxiliary counter read 20, and read off from the eight main counters the number of times those counters which paid off for him had been operated. The whole process takes about 20 to 30 sec. This procedure is simply a faster equivalent of playing the pinball machine 20 separate times, each time making an identical bet.

Subjects.—The *S*s for this experiment were five Johns Hopkins male undergraduates. They were naive about the purposes of this experiment, had not participated in gambling experiments before, and for the most part were relatively unfamiliar with gambling. The names of those who were invited to be *S*s were selected at random, but it was necessary to issue 23 invitations in order to get five acceptances. Therefore this cannot be regarded as a random sample of Johns Hopkins undergraduates. However, so far as the available information goes it

is a random sample of those Johns Hopkins undergraduates who are willing to participate in prolonged gambling experiments.

General procedure.—Each *S* was told that his purpose in the experiment was to make as much money as possible. He was also told that every game he would play in the experiment was mathematically "fair," that is, that in the long run he should break about even. However, he was told that the situation would be biased in his favor by $1.00 per session. This was accomplished by giving him $21 in chips at the beginning of each session and taking back only $20 in chips or cash at the end of each session. Each *S* knew that he was expected to pay any losses out of his own pocket, and that he would be paid his winnings in cash. However, when an *S* lost more money than he had available (one *S* lost $120 in one session), he was allowed to pay with an *IOU* until next time.

After all sessions were finished, each *S* was run for an extra session, in which the program of the gambling apparatus was prepared so as to cause him to win or lose enough so that he won an average of $1 per hour for the entire experiment. Each *S* spent a total of about 50 hr. in the experiment.

On all sessions in which the eight-counter apparatus was used, *S* was permitted to select which counters would be "his" and which would not. The counters which he chose as his would pay off in his favor; the others would be against him. He could not change selections in midsession. This arrangement was necessary because it was obvious that the eight-counter apparatus was programmed, and so it was necessary to make it clear that the program was not prepared unfairly. On the pinball machine, whose electrical connections and mode of operation were not visible, it was not necessary to do this.

Utility measurement sessions.—The first 12 sessions were devoted to the *N*-bets method of utility measurement. In Sessions 1–4, the probability of winning was 4/8 and the amount involved in each little bet was $0.10. In Sessions 5–8, the probability of winning was 4/8 and the amount involved in each little bet was $0.15. In Sessions 9–12, the probability of winning was 2/8 and the amount involved in each little bet was $0.10. The amounts of money whose utilities were measured in each session were plus $5.50, $4.50, $3.50, $2.50, $2.00, $1.50, $1.00, $0.90, $0.80 $0.70, $0.60, and $0.50; and minus those same amounts. For each amount of money, seven cards were prepared, each with a different value of *N*. A sample card read:

A. You have four numbers. If you roll any one of your numbers, you win $5.50. If you do not, you win nothing.

B. You have four numbers. You make 30 identical bets. On each bet, if you roll one of your numbers, you win $0.10. If you do not, you win nothing.

All other cards were similar, except that they included different amounts which could be won in the big bet and different numbers of small bets. To measure negative utilities, the verb was changed from "win" to "lose." In Sessions 5–8, each alternative began "You have two numbers." In Sessions 9–12, each small bet paid off $0.15 instead of $0.10.

In addition to the cards already discussed, other cards were prepared which were identical except that the value of *N* was left blank. This was necessary because it was possible that the indifference point for a given *S* might not fall, on one day, within the range of values of *N* which had been selected. Whenever that happened, *E* wrote in values of *N* which seemed more likely to be near the indifference point, and slipped the new cards into the pack. The *S*s were, of course, aware that *E* added cards to the pack from time to time, and they eventually learned that *E* kept on doing so until *S* had changed over from *A*-type alternatives to *B*-type alternatives, or vice versa. This knowledge would have made it possible for an *S* who was skilled at gambling to adopt the strategy of changing quickly on losing bets (thus maximizing the number of winning extra cards). To see whether anyone did this, counts were made of the number of extra cards used on winning and losing bets at various stages of the experiment. No indication could be found in these counts that any *S* used such a strategy. In fact, most *S*s tended to resist changing more stubbornly on losing bets than on winning ones.

When the approximate indifference point had been located in the first session, thereafter only cards reasonably near it were used in subsequent sessions. This reduced the total number of cards, and so the length of the sessions. It is probable that the selection of the range of cards presented to *S* had at least a minor effect on his indifference point. This effect could not be avoided, since some range had to be selected; it could only be hoped that the effect would be minor compared with the utility and subjective probability differences being studied. The effectiveness of the utility and subjective probability scales determined by this method in predicting choices among bets suggests that this hope was fulfilled.

Subjective probability measurement sessions.—Sessions 13–18 were subjective probability measurement sessions. The probabilities whose subjective probabilities were measured were 1/8, 2/8, 3/8, 4/8, 5/8, 6/8, and 7/8. Each subjective probability was determined twice in each session, once for positive *EV* bets and once for

negative *EV* bets. For each probability, cards offering that probability of winning various amounts of money compared with the certainty of winning $0.55 were prepared. A sample card read:

A. You have three numbers. If you roll any one of your numbers, you win $1.50. If you do not, you win nothing.
B. You receive $0.55 for sure.

Other cards used in determining the subjective probability of 3/8 were identical except for the amount of money which could be won in the *A*-alternative. Similar sets of cards were prepared for each other probability. As in the utility measurement sessions, extra cards with the amount of money in the *A*-alternative left blank were available, in case the indifference point fell outside the anticipated range.

Model-testing sessions.—Sessions 19–22 were the model-testing sessions. For the purpose of testing the model, the bets which have been extensively used in previous probability preference experiments were used again. Four lists of bets were used, with *EV*s of +$0.75, +$0.52½ and minus those same amounts. Table 1 gives those lists of bets. Each bet was paired with all others at its *EV* according to Ross's optimal order for pairs in the method of paired comparisons (12). The *S*s were required to choose one member of each pair. These bets were run off on the pinball machine. The *S*s made each choice four times, once in each session.

TABLE 1

BETS USED IN THE MODEL-TESTING SESSIONS

Positive Expected Value Bets

Probability	Payoff Cells	Amount Paid	
		$0.75 EV	*$0.52½ EV*
1/8	4	$6.00	$4.20
2/8	1, 7	3.00	2.10
3/8	2, 4, 6	2.00	1.40
4/8	2, 4, 7, 8	1.50	1.05
5/8	2, 3, 5, 7, 8	1.20	0.84
6/8	All but 3 & 6	1.00	0.70
7/8	All but 5	0.86	0.60
8/8	All	0.75	0.53

Negative Expected Value Bets

All negative expected value bets are identical with the corresponding positive expected value bets except that *S* loses the amount stated if the ball rolls into one of the payoff cells listed, instead of winning it.

Note.—The exact wording of these bets can be found in (5), p. 88.

RESULTS

Utility measurement sessions.—Regression lines were fitted to the results of each block of four utility measurement sessions for each *S* by the method of least squares, or a simpler approximation when little error resulted. The point at which *S* would presumably choose each alternative 50% of the time (the indifference point) was read from these lines. In three cases (out of 120) in which this procedure gave an obviously unreasonable 50% point, the actually observed 50% point was used instead.

To determine the utility functions, Equation 6 was solved for each amount of money used. The result of these calculations is three utility curves for each *S*. These curves are shown in Fig. 1. It is apparent that there are differences from one curve to another, but these differences do not seem to be systematic; they can probably be attributed to day-to-day variations in judgments. There is no evidence that the curve for Sessions 5–8, in which the $0.15 small bets were used, is consistently smaller for the high values of money than the other two curves, as it would be if the utility of *N* bets was less than *N* times the utility of one bet; or consistently larger, as it would be if the utility of *N* bets was greater than *N* times the utility of one bet. So it is reasonable to assume that this experiment remained within the sizes of *N* for which the *N*-bets assumption is acceptable. Since there does not seem to be any systematic difference between the different utility curves, the three utility values determined for each money value were averaged, and

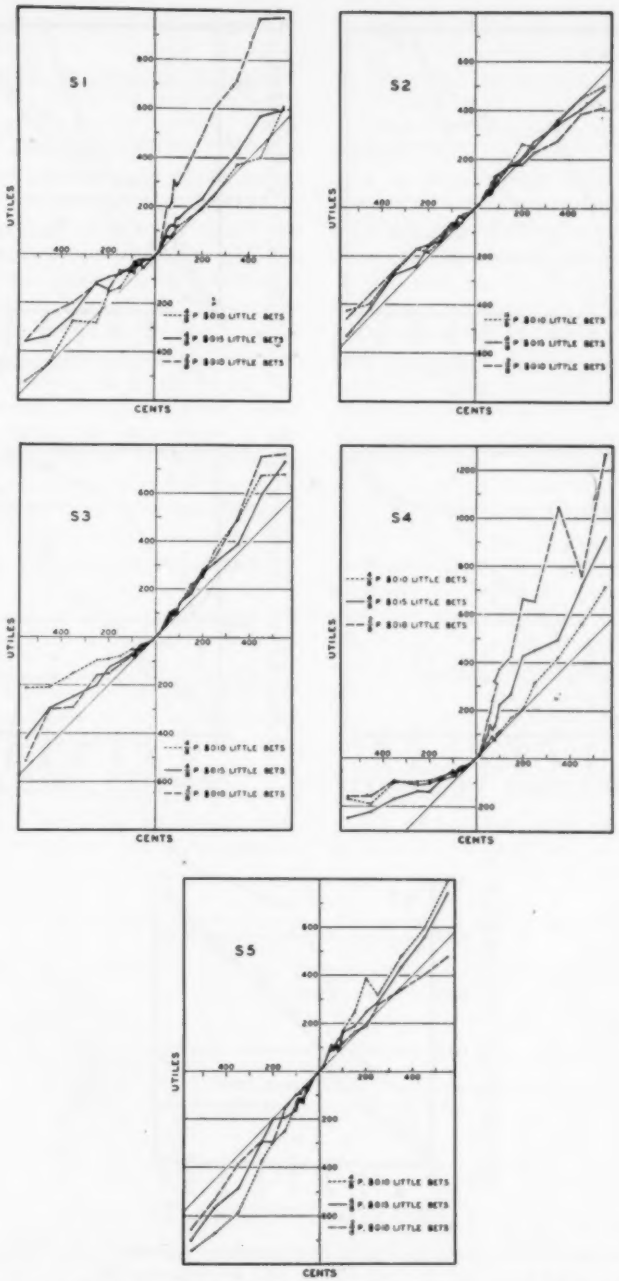

FIG. 1. Experimentally determined individual utility curves. The 45° line in each graph is the curve which would be obtained if the subjective value of money were equal to its objective value.

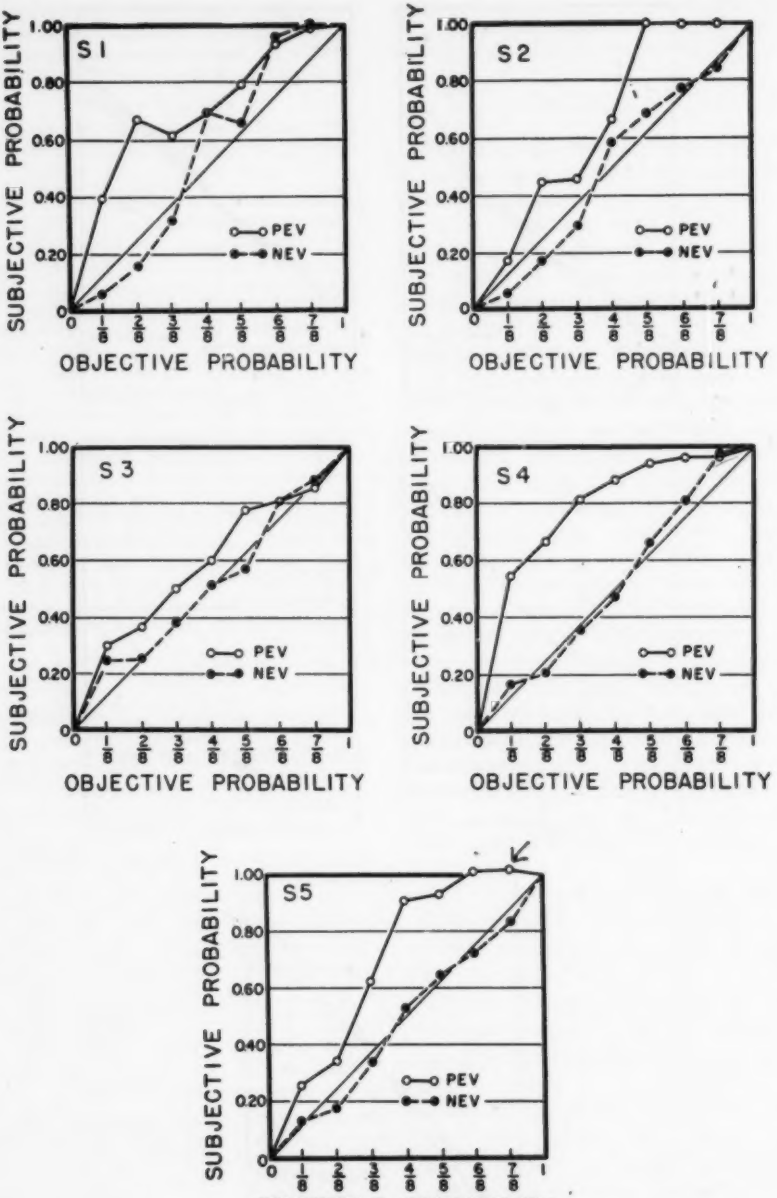

FIG. 2. Experimentally determined subjective probability functions. The functions labeled *PEV* are for winning bets. The functions labeled *NEV* are for losing bets. The 45° line is the function which would be obtained if subjective probability were equal to objective probability.

that average is used in all subsequent calculations.

The most outstanding fact about Fig. 1 is the relatively high degree of linearity of the utility curves. Most of them adhere fairly closely to the 45° line which represents the equation $u_{\$A} = \$A$ (where $A is measured in pennies). For those cases which deviate substantially from the 45° line, the deviation seems to be greater for the negative branch of the utility curve than for the positive branch. It should not be assumed that the utility of money is equal to its objective value for all possible values even for those *S*s who adhere to the 45° line most closely; instead, it seems likely that for large values of $A ($100 and more, for instance), the value of $u_{\$A}$ will be quite different from $A. However, most college students have had extensive experience with receiving and giving up amounts of money such as are used in this experiment. Their subjective impressions of the values of such sums of money are likely to be very close to the true values because of this long learning process. This generalization, if it is true enough often enough, offers a great simplification for future gambling experiments. If it is possible to use the physical value of money instead of its utility in the *SEU* maximization model, then application of the model will be much easier.

Subjective probability measurement sessions.—Graphs similar to those used in determining 50% points for the utility measurement sessions were used for the subjective probability measurement sessions, except that the *X*-axis contained amount of money rather than value of *N*. Mid-points were determined by the same methods used for the utility measurement sessions. Then equations like Equation 8 were solved for each positive and negative subjective probability. Graphs of the resulting subjective probability functions are shown in Fig. 2. The most outstanding fact about Fig. 2 is that in general *S*s tend to overestimate their chances of winning money, while retaining a fairly objective estimate of their chances of losing it. This result agrees pretty well with casual observation and common sense. Notice that one subjective probability function goes slightly above 1. This finding, though surprising, is permitted by the *SEU* and *SEM* models.

The test of the SEU maximization hypothesis.—The choices made during

TABLE 2

NUMBER OF CORRECT AND INCORRECT PREDICTIONS BASED ON THE *SEU* MAXIMIZATION HYPOTHESIS

Expected Value Level	Subject										Mean	
	1		2		3		4		5			
	Right	Wrong	Right	Wrong	Right	Wrong	Right	Wrong	Right	Wrong	Right	Wrong
+$0.52–1/2	100	12	71	41	99	13	107	5	72	40	89.8	22.2
+$0.75	81	31	73	39	100	12	105	7	93	19	90.4	21.6
−$0.52–1/2	78	34	71	41	61	51	95	17	66	46	74.2	37.8
−$0.75	64	48	87	25	45	67	91	21	70	42	71.4	40.6

Note.—By chance 56 correct predictions should be expected in each group. By chi square, 67 predictions is significantly different from 56 at the .05 level; 70 at the .01 level; and 74 at the .001 level. Accordingly, all predictions are significantly better than chance except those for *S*s 1 and 3 on the − $0.75 *EV* level and for *S*s 3 and 5 at the − $0.52–1/2 *EV* level.

the model-testing sessions, Sessions 19–22, are not interesting in themselves. They were very similar to the usual results of probability preference experiments (**3, 5**). The interesting thing about these choices is whether or not they correspond to the predictions made by the *SEU* model. Table 2 shows the number of correct and incorrect predictions for each of the four *EV* levels for each *S*. It is at once apparent that the model predicts significantly better than chance (i.e., predictions from the model are better than those that could be made by flipping a coin). It is also apparent that the predictions from the model include a number of errors. These errors are of two kinds. First, there are cases in which *S*s chose one way on a given pair in some of Sessions 19–22 and the other way on other sessions. Since the model predicts 100% choice of the alternative with the higher *SEU*, any such day-to-day variation in choices results in erroneous predictions. Such errors do little damage to the model. The other and more serious kind of error occurs when *S* chooses consistently in all four model-testing sessions, but the *SEU* model predicts the opposite of the choices which actually occur. On the average, each *S* made 18.8 errors of the first kind and only 10.6 errors of the second kind at each *EV* level. (Note that this means that only an average of 2.65 pairs per *S* per *EV* level were completely wrongly predicted; each completely wrong prediction produces four errors.) An analysis of variance of the errors shows that this difference is very highly significant. Variance due to *S*s, of course, is also significant. And the difference between errors on positive *EV* and negative *EV* bets is significant. This last variable interacts with *S*s. The nature of the effect, as can be seen from inspecting Table 2, is that errors are in general more frequent on negative *EV* than on positive *EV* bets, but this is not true for some *S*s. It is interesting that neither *S*s nor positive *EV* bets vs. negative *EV* bets interacts significantly with consistent vs. inconsistent errors.

The next task is to account for the wrong predictions. Some of them, as we have seen, result from inconsistent choices in the model-testing sessions. Others may perhaps be accounted for by the size of the *SEU* differences involved. An obvious hypothesis is that *S*s are more likely to choose consistently with prediction when the two bets between which they are choosing are far apart from one another in *SEU*. This hypothesis is strengthened by the fact that the median *SEU* difference between pairs of bets is much smaller for the negative *EV* bets for all *S*s than it is for the positive *EV* bets. This is consistent with the fact that, in general, prediction is less satisfactory on the negative *EV* bets. In order to test this hypothesis more precisely, product-moment correlation coefficients were calculated between the size of the *SEU* difference between each pair of bets and the number of choices consistent with prediction on that pair. These correlations are small (median value: + .23), but all of them are positive. The reason they are small is that the number of choices consistent with prediction can have only five values, 0, 1, 2, 3, and 4. When a variable which can assume only a few values is correlated with some other variable, the correlation coefficient is inevitably low. However, a simple sign test shows that the probability of obtaining 20 correlation coefficients with the same sign, if the true correlation is 0, is $2\ (1/2)^{20}$—

TABLE 3

NUMBER OF CORRECT AND INCORRECT PREDICTIONS BASED ON THE *EU* MAXIMIZATION HYPOTHESIS

Expected Value Level	Subject										Mean	
	1		2		3		4		5			
	Right	Wrong	Right	Wrong	Right	Wrong	Right	Wrong	Right	Wrong	Right	Wrong
+$0.52–1/2	12	100	73	39	41	71	37	75	20	92	36.6	75.4
+$0.75	19	93	59	53	40	72	25	87	41	71	36.8	75.2
−$0.52–1/2	38	74	57	55	39	73	103	9	66	46	60.6	51.4
−$0.75	24	88	63	49	47	65	99	13	82	30	63.0	49.0

very small indeed. This confirms the hypothesis that there is a positive relationship between size of *SEU* difference and number of choices consistent with prediction.

A test of the EU maximization hypothesis.—The next question to ask is: How well can the utilities alone and the subjective probabilities alone predict choices?

To test the predictive value of the utilities considered alone is essentially equivalent to testing the *EU* maximization model proposed by Bernoulli and von Neumann and Morgenstern. In order to do it, all that is needed is to calculate the *EU* for each *S* for each bet, using Equation 2. This was done. The resulting predictions were compared with the choices made in the model-testing sessions. The results of this comparison are shown in Table 3. It is at once evident that the *EU* maximization hypothesis does not predict better than chance. Indeed, if an analysis of variance of the error scores is made and the error estimate from that analysis is used to estimate the standard error of the mean, then the *EU* maximization hypothesis predicts significantly ($P = .05$) worse than chance on the positive *EV* bets. (The results of that analysis of variance, incidentally, show that *S*s, the difference between positive *EV* and negative *EV* bets, and their interaction significantly affect the number of errors.) None of the various other comparisons which were reported for the *SEU* model offers any evidence, when made on the *EU* model, that the *EU* model is any good. So far as the evidence of this experiment goes, it is necessary to conclude that the *EU* maximization model fails to predict simple gambling decisions.

The reason for the failure of the *EU* maximization model is, presumably, the near-linearity of the utility scales used. If the utility scales are actually relatively linear, then any predictions of choices between constant-*EV* bets must be made on the basis of minor irregularities in the curves, and it is not surprising that such predictions fail to work out. The only successful predictions based on *EU* values, *S* 4's *NEV* bets, are consistent with this argument. This *S*'s negative utility curve deviates substantially from the $u_{\$A} = \$A$ line, and his negative subjective probability curve deviates very little from the $P^* = p$ line.

The relative linearity of the utility curves, if it is a general phenomenon, means that the *EU* maximization model will never, as a practical matter, be very useful in gambling experiments. This does not, of course,

preclude the possibility that it might be useful in other situations, particularly those which use larger sums of money. However, the evidence does strongly suggest that whenever the *EU* maximization model is useful, the *SEU* maximization model will be more useful.

A test of the subjectively expected money (SEM) maximization model.— We have examined the *SEU* and *EU* maximization models. It remains to examine the *SEM* maximization model, in which prediction of choices is affected only by subjective probabilities. In the original calculation of the subjective probability scales given in Fig. 2, utility values were used. If the predictive value of the *SEM* maximization model is to be tested, it is not permissible to use utilities in the original calculation of the subjective probabilities which go into the model. Consequently, all subjective probabilities were recalculated using money values instead of utility values in equations like Equation 8. The resulting subjective probability functions differ only in minor ways from those given in Fig. 2.

For each bet used in the model-testing sessions, an *SEM* value was calculated from the new subjective probability functions. Then the predictions based on the hypothesis of *SEM* maximization were compared with the choices in the model-testing session. Table 4 shows the results. It is at once apparent that the *SEM* maximization model predicts quite well. Comparison of Table 4 with Table 2 shows that the *SEM* maximization model predicts nonsignificantly better than the *SEU* maximization model. Analysis of variance of the *SEM* model errors gives essentially the same picture as that for the *SEU* model. The same generalizations about errors and about the superiority of positive *EV* predictions over negative *EV* predictions which were made for the *SEU* model apply to the *SEM* model also. Indeed, the only respect in which the *SEM* model is inferior to the *SEU* model is that when the correlations between *SEM* difference and number of choices consistent with prediction are calculated, only 17 instead of 20 of the correlation coefficients come out positive. The probability of getting so many correlation coefficients of the same sign if the true correlation is 0 is .0026.

It must be concluded that the *SEM*

TABLE 4

NUMBER OF CORRECT AND INCORRECT PREDICTIONS BASED ON THE *SEM* MAXIMIZATION HYPOTHESIS

Expected Value Level	Subject										Mean	
	1		2		3		4		5			
	Right	Wrong	Right	Wrong	Right	Wrong	Right	Wrong	Right	Wrong	Right	Wrong
+$0.52–1/2	100	12	71	41	99	13	107	5	72	40	93.4	18.6
+$0.75	93	19	73	39	100	12	105	7	93	19	92.4	19.6
−$0.52–1/2	78	34	69	43	61	51	93	19	66	46	74.2	37.8
−$0.75	76	36	75	37	45	67	91	21	70	42	72.2	39.8

Note.—By chance 56 correct predictions should be expected in each group. By chi square, 67 predictions is significantly different from 56 at the 0.05 level; 70 at the 0.01 level; and 74 at the 0.001 level. Accordingly, all predictions are significantly better than chance except for *S* 3 at both negative *EV* levels and *S* 5 at the − $0.52–1/2 level.

model predicts choices just as well as the *SEU* model does. Indeed, examination of the predictions shows that the *SEM* model and the *SEU* model usually predict the same or almost the same rank order for the bets. The reason why is that the *SEM* values correlate highly with the *SEU* values. The median correlation is .96; the lowest is .82.

Discussion

If the results of this experiment are taken literally, they mean the following things. First, over the range of money used in this experiment (+ $5.50 to − $5.50) the utility of money is fairly linearly related to its dollar value for most *S*s. Secondly, the *SEU* maximization model predicts choices among simple bets rather well. Thirdly, there is a positive (though not perfect) correlation between the size of the *SEU* difference between two bets and the probability that the *S* will choose the one with the larger *SEU*. Fourthly, the *EU* maximization model is no good at predicting choices among simple bets like those used in this experiment. This is presumably because utility is fairly linear with money value over the ranges used in this experiment, and so the sizes of the *EU* differences among the bets used in the model-testing sessions of this experiment are very small. Finally, a model which asserts that *S*s choose so as to maximize *SEM* predicts just as well as the *SEU* maximization model, and for the most part makes the same predictions. This fact can be explained, of course, by the fact that the utility curves are approximately linear, so that *SEM* and *SEU* values for any bet are very highly correlated with one another. Most of these conclusions depend for their validity on the assumption that the *N*-bets method is a satisfactory method of utility measurement. The internal consistencies in the results which indicate its acceptability have already been discussed.

The results of this experiment are entirely consistent with the results of the various probability preference experiments. They suggest that the most important determiner of choices among bets is subjective probabilities, and that bets with the same subjective probability ought to be preferred to about the same extent when compared with others at the same *EV* level. That is just what happened in one probability preference experiment (**5**).

Generalizations from this experiment should be made cautiously. First, neither the *SEU* model nor this experiment offers any way of predicting choices among bets which include possibilities both of winning and of losing. The *SEU* values of the winning possibilities and of the losing possibilities can be algebraically added to get a composite *SEU*, but whether such composite *SEU* values will predict choices is an experimental problem. Secondly, the *SEU* model has no true zero point, and so does not predict whether a given bet will be accepted or refused, if the alternative is simply not gambling. Finally, there is no guarantee that subjective probabilities are independent of the utilities with which they are associated, though the *SEU* model assumes that they are. With very large bets or very small bets, the form of the subjective probability function might be quite different from that found in an experiment like this one for a given *S*. All of these questions can be settled by appropriate experiments.

More theoretical work needs to be done on the *SEU* model. Ideally, it should be a probabilistic model, predicting the probability of choice of the alternative with the higher *SEU* as a function of the size of the *SEU* difference. The obvious form for a theoretical curve relating probability of choice of alternative *A* over alternative *B* to size of $SEU_A - SEU_B$ is a normal ogive, with the 50% point at a *SEU* difference of 0. But other intuitively plausible curves can also be conceived. At any rate, until better methods of measuring utilities and subjective probabilities are found, there is little reason for going into such theo-

retical refinements, because there is no hope of testing them experimentally. The improvement of methods of utility and subjective probability measurement is still the most important task in the study of gambling behavior.

Summary

A very simple mathematical model was developed for predicting choices among bets. This model, based on the concepts of subjective value or utility of money and subjective probability, asserts that *S*s choose the bet with the maximum subjectively expected utility. An experiment was designed to test this model. The utility of money was measured by assuming that the utility of N identical bets is N times the utility of one such bet—an assumption which the results of the experiment verify. Subjective probabilities were measured by other choices. Finally, still other choices were used to test the model. The model predicted substantially better than chance. Another model, which assumes that *S*s choose so as to maximize expected utility, failed to predict choices successfully. A third model, which uses only the concept of subjective probability, predicted as well as the subjectively expected utility maximization model.

It is concluded that the subjectively expected utility maximization model is adequate to account for the results of this experiment, and that subjective probabilities are much more important than utilities in determining choices among bets such as those used in this experiment.

References

1. Bernoulli, D. Specimen theoriae novae de mensura sortis. *Comentarii Academiae Scientiarum Imperiales Petropolitanae*, 1738, **5**, 175–192. (Translated by L. Sommer in *Econometrica*, 1954, **22**, 23–36.)
2. Coombs, C. H., & Beardslee, D. C. On decision-making under uncertainty. In R. M. Thrall, C. H. Coombs, & R. L. Davis (Eds.) *Decision processes*. New York: Wiley, 1954.
3. Edwards, W. Probability-preferences in gambling. *Amer. J. Psychol.*, 1953, **66**, 349–364.
4. Edwards, W. Probability-preferences among bets with differing expected values. *Amer. J. Psychol.*, 1954, **67**, 56–67.
5. Edwards, W. The reliability of probability-preferences. *Amer. J. Psychol.*, 1954, **67**, 68–95.
6. Edwards, W. The theory of decision making. *Psychol. Bull.*, 1954, **51**, 380–417.
7. Edwards, W. Variance preferences in gambling. *Amer. J. Psychol.*, 1954, **67**, 441–452.
8. Edwards, W. An attempt to predict gambling decisions. In *Mathematical models of human behavior—Proceedings of a symposium*. Stamford, Conn.: Dunlap & Associates, 1955.
9. Markowitz, H. The utility of wealth. *J. polit. Econ.*, 1952, **60**, 151–158.
10. Mosteller, F., & Nogee, P. An experimental measurement of utility. *J. polit. Econ.*, 1951, **59**, 371–404.
11. Preston, M. G., & Baratta, P. An experimental study of the auction-value of an uncertain outcome. *Amer. J. Psychol.*, 1948, **61**, 183–193.
12. Ross, R. T. Optimum orders for the presentation of pairs in the method of paired comparisons. *J. educ. Psychol.*, 1934, **25**, 375–382.
13. von Neumann, J., & Morgenstern, O. *Theory of games and economic behavior*. (2nd Ed.) Princeton: Princeton Univ. Press, 1947.

(Received November 1, 1954)

Journal of Experimental Psychology
Vol. 50, No. 3, 1955

VARIATIONS IN KNOWLEDGE OF COMPONENT PERFORMANCE AND ITS EFFECTS UPON PART–PART AND PART–WHOLE RELATIONS [1]

EDWARD A. BILODEAU

Skill Components Research Laboratory, AF Personnel and Training Research Center

A previous article (1) discussed some of the possible contributions of component activities to the total psychomotor task and suggested some possible means of examining the relations between components of a task and the total task. That article should be considered as the introduction to the experiments reported here. The previous paper dealt with the analysis of simultaneous and nonsimultaneous occurrences of two or more responses in fairly complex psychomotor performance situations. Data were presented on two tasks which showed that means and variances for criterion combinations of on-target scores could be fairly accurately predicted by assuming that the times of occurrence of scoring in one response dimension are independent of the times of scoring in a second dimension. This can be expressed more precisely as follows,

$$A_o \times B_o = AB_p = AB_o, \quad (1)$$

where A and B represent two different response dimensions, AB represents the combination score or the proportion of time on target in both dimensions simultaneously, and o and p identify proportions (probabilities) of time on target as obtained and predicted, respectively.

[1] This research was carried out under the Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas, in support of Project 7707. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government.

The hypothesis of independence among component behaviors was tested with the Pedestal Sight Manipulation Test (2) and the Stevens Pursuitmeter (3). The results showed a slight tendency for a positive relation between two components and a zero relationship if a third component were added (ABC). At any rate, no consequential error (in a practical sense) was made in predicting AB proficiency whenever independence was assumed.[2]

An implication of the chance relationship is that redistributions of

[2] Unfortunately, statistical theory does not yet appear sufficiently advanced for adequate tests of significance. The product of two independent variables, each of which is normally distributed, is not normally distributed. Further, the distribution of a product can be skewed. Since the sampling distribution of the product is unknown, it is hazardous to compare observed and predicted means or variances.

Because of the difficulties over statistical comparisons of the data, the question of the desirability of collecting data at all is raised. This question is not easily laid to rest. The author believes that a contribution can be made without making a statistical evaluation of the significance of the difference between means. For example, if it can be said that $\overline{AB}_o - \overline{AB}_p = 2$ percentage points or less for several psychomotor devices, and for extended periods of practice, the statement appears worthwhile without statistical evaluation. At least, another investigator could expect to get similar discrepancies under similar circumstances and could probably assume that Formula 1 will lead to values of $\overline{AB}_p$ which are within tolerable or practical ranges of $\overline{AB}_o$. More sophisticated evaluations of the data are, of course, desirable, and it is hoped that a treatment of the distribution of products will be made available.

part proficiencies will alter the magnitude of the combination score in specifiable ways. It was also stressed that the alteration of the combination score can perhaps be attained without changing *S*'s fundamental ability to perform the parts of the total task (when each part is practiced separately). As an example of this point, suppose an *S* practicing two parts simultaneously scores $A_o = .90$ and $B_o = .20$. For this distribution of part proficiencies, a proportion of .18 should be obtained as a criterion of proficiency at remaining on target in *A* and *B* simultaneously. Suppose that the same *S*, however, could score $A_o = .80$ and $B_o = .30$. In this case, the predicted *AB* score is .24 or 6 points higher than with the previous distribution, but the sum of the individual probabilities, *A* and *B*, is not changed. If it can be experimentally demonstrated that *S* can alternate from one distribution to the other from trial to trial, then it might be allowed that *S* is exhibiting a constant skill level, yet systematically varying his criterion score.

The previous paper outlined two methods by which part proficiencies might be alterable when all parts are simultaneously introduced, without changing *S*'s ability to perform each of the parts when performed as parts; both methods are used in the present study. In the first method, *S*'s attention to environmental stimuli (*A* and *B* display needles) is redistributed through manipulation of knowledge of results (KR); this is done by periodically informing *S* which of his part proficiencies is poorest. In the second method, *S*'s attention is redistributed in a manner procedurally different from the first; this is done by instructing *S* to pay twice as much attention to one particular part of the task.

The study outlined above must allow the individual *S* a relatively long period of practice because the task must be difficult and it must be assumed that *S*'s performance of the parts is stable. Thus, the experiment is necessarily lengthy. The study is broken down into two experiments. Basically, the first experiment represents but an extended training period for assuring stable and proficient performance. The extended training given in Exp. 1, however, also afforded opportunity for evaluating the effects of adding certain kinds of KR both early and late in training. In Exp. 2, the *S*s were subjected to extreme conditions calculated to alter the usual modes of attending to a multidimensional task in order to examine part-part and part-whole relations.

Method

Apparatus.—Two units of the SAM Multidimensional Pursuit Test (CM813E) were used. This apparatus is described by Melton as follows: ". . . consists of an instrument panel and simulated airplane controls. During a test trial, the various pointers of the panel meters drift about in an irregular and continuously varying pattern. The subject's task is to manipulate the controls so that the meter pointers are kept at specified points on their respective scales. The 'stick' controls the 'climb' and 'bank' indicators, the rudder pedals control a 'turn' indicator, and the throttle governs the position of the needle on an 'airspeed' indicator.

"Movement of the various needles on the instrument panel is governed by a motor which turns four cams, spaced at intervals on one shaft. Each cam controls the position of the sliding contact on a wire-wound potentiometer. Corresponding to each of the motor driven potentiometers there are four potentiometers which are connected to the various airplane controls. When one of the motor driven potentiometers shifts a needle from its specified point on the scale, a compensatory adjustment of the appropriate airplane control will bring the needle back to its origin" (3, pp. 7–18).

The basic apparatus above was modified to present a simpler display panel. Indicator scales of simple and standard design were substituted for the original scales. The modified

TABLE 1

DISPLAY–CONTROL RELATIONSHIPS FOR THE MULTIDIMENSIONAL PURSUIT TEST

Control	Control Motion	Indicator	Indicator Effect
Stick (left hand)	forward	air speed (S)	right
	backward	air speed (S)	left
Pedals (feet)	right	heading (H)	right
	left	heading (H)	left
Stick (right hand)	forward	altitude (A)	right
	backward	altitude (A)	left
Stick (right hand)	right	bank (B)	right
	left	bank (B)	left

scales represented zero at the center and deviations from zero by markers spaced at uniform intervals to left and right of zero. The *S*'s task, then, can be described as attempting to zero four indicator needles in a compensatory manner by appropriate manipulations of four controls. The display-control relationships are outlined in Table 1. For simplicity, the four controls or components are called S, H, A, B.

Procedure in Exp. 1

The variable was the kind of KR administered to 72 naive airmen from Lackland AFB. Three different varieties of KR were evaluated during a 20-day period of practice. After 60-sec. trials of practice, the *S*s of Treatment LC (lowest component) were told which of their four components ranked lowest; the *S*s of Treatment RC (random component) were told "which of their four components ranked lowest," but the component was actually selected by *E* on a random basis, and *S*s of Treatment NC (no component) were given no extra KR by *E*. Treatment RC was included as a test of the effects of the mere identification of a component for *S*, as opposed to correct identification of the worst component. Treatment RC, then, was used as a sort of motivational control. Treatment NC represents a standard manner of administering a psychomotor test without giving KR beyond that which *S* can see for himself, and Treatment LC represents a not uncommon manner of providing a certain amount of extra KR, although by no means the only way.

By using different treatment sequences for four groups, it was possible to arrange for a test of the effects of KR when introduced immediately or relatively late within the practice session. The sequence of treatments for the four groups is presented in Table 2.

Standard Electric timers measured the cumulative time that the controls were within the target extent definitions. Besides timers for each of the four components, three additional timers usually took measures of AB, HAB, and SHAB. AB, for example, represents the time the controls were correctly positioned within the limits of A and B simultaneously. The time-on-target scores were recorded after each 60-sec. practice trial and all time scores in this paper are presented as percentages.

Although *S*s were not so instructed, the on-off target limits for each of the four indicator needles were two units to the right and left of the zero point (deviation of any needle from zero by as much as, say, .25 units was clearly discriminable). The instructions emphasized that the task was to center, and to keep centered as much as possible, all of the needles. It is unlikely that the *S*s identified the scoring limits.

Additional points of procedure were as follows: the number of trials per day was 8, rest between trials was 60 sec., number of *S*s per group was 18, and number of days of practice was 20.

On Days 2, 6, 10, 14, and 18, a multichannel operations recorder was used to record various aspects of momentary on-off target performance. The clocks and recorder were wired in parallel so that the readings of one instrument could be checked against the other. Records were made of S, H, A, B, and SHAB on the last 7 trials, of AB and HAB on Trials 2, 3, 7, and 8, of SA and SB on Trial 4, of AH and BH on Trial 5, and of AB and BH on Trial 6. These records made possible an analysis of time of occurrence and duration of many on-off target events at selected points during the learning session. It was not convenient to obtain a record of all possible combinations at any given moment of practice.

Procedure in Exp. 2

Experiment 2 began on Day 21 of practice because proficient *S*s were required. The 72 *S*s of Exp. 1 were reassembled into two matched groups and run an additional four days in a manner identical with that of the preceding 20 days except as noted below.

TABLE 2

EXPERIMENTAL DESIGN, EXP. 1

Group	Blocks of Days		
	1–12	13–16	17–20
LC (Lowest of 4 components)	LC	LC	LC
RC (Random one of 4 components)	RC	RC	LC
NC (No component reported lowest)	NC	LC	LC
NC (No component reported lowest)	NC	RC	LC

Group C (control).—After each trial, a randomly selected component was identified as lowest. The *S*s were instructed that these scores represented their worst component. (Same as RC, Exp. 1.)

Group SI (special instruction).—These *S*s were given their lowest component after each trial and told, "From now on I want you to pay twice as much attention to your A needle. Watch the A needle twice as often as you do any other needle *and* be sure that you make about twice as many corrections on the A control as any other control. I'll tell you what your lowest score is after each trial."

Another important difference between Exp. 1 and 2 is that the scoring band for A was reduced by one-half (to 1 marker right and left of zero) for both groups. It was, however, almost impossible for any *S* to deduce this change in calibration. The *S*s of Group SI, of course, had the advantage of being instructed to attend twice as often to A (needle and control) as to any other needle and control. A further advantage was that whenever A ranked lowest, A was reported as lowest. Thus, two variables were purposely confounded in Group SI. Both variables should act in the same direction and should lead to better A performance for Group SI. The reason for using two variables for this group was that substantial differences between component performances from group to group were desired. There was no intention of parceling out the separate effects of instructions (to pay twice as much attention to A) and identifying the lowest component. Experiment 2, then, represents extreme conditions calculated to disrupt the usual modes of responding to a four-component complex.[3]

For the last day of practice (Day 24) additional change in the behavior of Group SI was sought by arbitrarily reporting B, after each trial, to be the lowest component of the four. The *S*s were then faced with the problem of attending twice as often to A as to any other component, while acknowledging that their worst performance was with respect to B.

[3] Halving the scoring area does not alter *S* behavior, but does alter the magnitude of the A scores obtained. This procedure results in a situation where one component of a complex is more important or critical than the others, for it alters the weighting of A in the combination scores.

Halving A should result in A's ranking fourth provided no special instructions for attending to A have been administered. Actually, A had an average rank of 2.2 on Trial 1 of Day 21 for Group SI, demonstrating the immediate effect of the special instructions. On later trials, if any S's attention to A tended to lapse, the LC procedure could be expected to augment A.

TABLE 3

MEAN DIFFERENCES BETWEEN TRIALS 2 AND 3 IN PERCENTAGE POINTS ON DAY 2*

Group	Component				Sum C^+	Sum C^-
	$\overline{S}$	$\overline{H}$	$\overline{A}$	$\overline{B}$		
LC	+2.6	−2.2	+2.1	+4.1	+9.1	−.9
RC	+7.7	−1.5	+1.1	+1.2	+4.4	+1.4

* Sum C^+ represents the mean total gain for any of four components identified for *S* as the worst component. Sum C^- represents the average gain for the three components not represented to *S* as the worst component.

RESULTS

It is not possible to present all the data nor to present all the comparisons which might be of interest. For example, over 96,000 time scores were recorded. General tabular summaries of the data have been deposited with ADI.[4]

Experiment 1

Examination of the daily means of components and combinations over Days 1–20 led to the conclusion that there was little, if any, difference between groups. However, even casual observation of other aspects of the data showed promise of differences when, say, Trial $n + 1$ is examined in the light of which component was reported as lowest on Trial n. A thorough trial-by-trial analysis, though of general interest, is omitted from the present report because of space limitations.

A cursory examination of this topic may be made, however, through the entries of Table 3. Group LC shows

[4] Tables from which the figures were constructed have been deposited with the American Documentation Institute. Order Document No. 4608 remitting $1.25 for microfilm or $1.25 for photocopies.

a mean increase of 9.1 percentage points for the component represented to *S* as being his worst. On the other hand, the average gain of the other three components is negative (− .9) and substantial positive gains could otherwise be expected. This suggests that the gain for the worst component is at the expense of the three remaining components since the gain for any component tends to be associated with an LC treatment. The other entries represent the average gain for each component regardless of the KR schedule. A similar calculation for Group RC is represented in the bottom row; the conclusion here is much the same as for Group LC, except that component S is high for some unknown reason.

Each of the seven major performance measures increased in magnitude as a function of day of practice. The grand means for all *S*s for Days 1 and 20 in Table 4 are suggestive of the level of performance and amount of improvement taking place. Actually, peak performance was reached between Days 17 and 20 when the seven performance measures are separately examined over all days. It can be rather safely concluded that 20 days of practice was about enough to bring each measure to, or very close to, its maximum (or plateau).

Analysis of component independence.—Obtained and predicted means for time on target in AB, HAB, and SHAB are presented in Table 5, where the analysis through Day 18 is of immediate concern. $\overline{AB}_p$ consistently underestimates $\overline{AB}_o$ by about 4 percentage points for each of the five day samples, suggesting the relationship between A and B is probably (*a*) positive and small, and (*b*) invariant with practice. A comparison of $\overline{HAB}_p$ and $\overline{HAB}_o$ leads to the same suggestion except that the absolute discrepancy is smaller, i.e., circa 2 percentage points. In the comparison of $\overline{SHAB}_p$ with $\overline{SHAB}_o$, the discrepancy is even less, but there is a very slight suggestion that the predicted mean overestimates the one obtained early in practice, but underestimates it late in practice. This is the only evidence of a changing dependency as a function of practice.

The analysis thus far has emphasized A and B, where but one control is operated in two dimensions. Comparisons across controls for pairs AH, BH, AS, and BS were also made. The striking feature of all comparisons was that the predicted mean was

TABLE 4

GRAND MEAN SCORES FOR DAYS 1 AND 20 FOR SELECTED RESPONSE MEASURES

Day	Measure						
	$\overline{S}$	$\overline{H}$	$\overline{A}$	$\overline{B}$	$\overline{AB}$	$\overline{HAB}$	$\overline{SHAB}$
1	51.1	53 2	53.5	57.5	33.6	17.2	8.6
20	73.7	74.1	71.6	78.8*	59.6*	44.3*	32.8*

* n = 71; otherwise n = 72.

TABLE 5

OBTAINED AND PREDICTED MEANS ON TRIAL 2 FOR SELECTED DAYS FOR ALL GROUPS COMBINED AND FOR GROUPS C AND SI

Day	$\overline{AB}_o$	$\overline{AB}_p$	$\overline{HAB}_o$	$\overline{HAB}_p$	$\overline{SHAB}_o$	$\overline{SHAB}_p$
Experiment 1						
2	39.0	35.1	22.1	21.0	9.7	11.4
6	50.2	46.1	33.3	30.9	19.9	20.0
10	54.9	51.3	36.7	34.5	23.6	23.3
14	55.2	51.6	37.8	35.5	25.1	24.7
18	59.9	56.3	43.2	40.9	30.5	29.4
Experiment 2						
22 (all)	43.9	41.5	29.0	28.5	17.6	18.2
22 C	41.8	38.7	30.7	29.5	20.3	20.9
22 SI	46.0	44.4	27.3	27.5	15.0	15.6

always very close to the one obtained. It appears that the greatest difference between the obtained and predicted scores was with respect to AB. The graphic recordings were used to pursue this lead further.

The relationship of A and B.—The graphical records for Days 2 and 18 were scored for the four possible events with respect to scoring in A and B. All 72 *S*s were scored for Trial 2, the examination of the record being made at each of 60 1-sec. intervals during the trial. The 72 2×2 contingency tables were analyzed by groups, the hypothesis of independent events being used to calculate predicted frequencies. All four χ^2's were significant beyond the .05 level. Of the 72 tables, 60 showed obtained AB frequencies greater than those predicted, and 12 obtained frequencies were smaller than those predicted. The latter were fairly uniformly distributed over the groups.

For Day 18, as compared with Day 2, the χ^2's were generally larger. The χ^2 for each of the four groups was significant beyond the .01 level, but the mean discrepancy between predicted and obtained AB values decreased from 3.4 (Day 2) to 2.6 points (Day 18), indicating a greater than chance dependency, yet strongly arguing against an increasing strength of dependency as a function of practice. Further evidence to support this contention was that 16 of the 72 analyses showed a negative dependence between A and B (compared with 12 on Day 2). A Pearson r was computed between the discrepancies ($\overline{AB}_o - \overline{AB}_p$) from a chance relationship for Days 2 and 18. This r was + .05 and indicates that the individual differences as regards departures from chance are not strong, to say the least.

TABLE 6

EXPERIMENT 2: DAILY MEANS FOR EACH COMPONENT OF PRACTICE FOR GROUPS C AND SI*

Group	Measure	Days			
		21	22	23	24
C					
	$\bar{S}$	72.8	73.7	71.2	74.4
	$\bar{H}$	72.6	73.6	73.4	74.4
	$\bar{A}$	48.0	45.3	48.8	48.4
	$\bar{B}$	79.5	79.3	80.7	79.3
	$\bar{A}\bar{B}$	41.3		42.0	40.9
	$\bar{H}A\bar{B}$	29.9		30.3	30.3
	$\bar{S}HA\bar{B}$	22.3		22.1	23.8
	Σ ($\bar{S}$,$\bar{H}$,$\bar{A}$,$\bar{B}$)	273	272	274	277
	Σ ($\bar{A}$,B)	128	125	130	128
SI					
	$\bar{S}$	63.0	62.2	61.3	61.4
	$\bar{H}$	63.2	60.9	63.1	60.2
	$\bar{A}$	68.0	62.8	66.7	61.8
	$\bar{B}$	71.1	70.8	71.3	78.6
	$\bar{A}\bar{B}$	49.8		49.2	50.0
	$\bar{H}A\bar{B}$	30.9		30.2	29.1
	$\bar{S}HA\bar{B}$	20.4		18.6	18.9
	Σ ($\bar{S}$,$\bar{H}$,A,$\bar{B}$)	265	257	262	262
	Σ ($\bar{A}$,$\bar{B}$)	139	134	138	140

* $n = 36$.

Experiment 2

The mean percentages of time on target for Groups C and SI are given in Table 6. The effect of reducing the scoring area of A by one-half is immediately apparent in the scores of Group C when the level of A is compared with the levels of S, H, and B. From Day 20 to Day 21, A has fallen approximately 24 percentage points to rank a very poor fourth on Day 21. By contrast, the A score for Group SI ranks second and has fallen from Day 20 to 21 about 4 percentage points. Group SI, however, shows a loss in all components. The losses from Day 20 to Day 21 are shown below.

Group	S	H	A	B	AB
C	−1	−1	−24	0	−19
SI	−11	−11	−4	−8	−10

From Day 20 to Day 21, Group C has lost 19 points in AB and Group SI has lost 10 points. Apparently, instructing the *S*s of Group SI to pay twice as much attention to A has resulted in about one-half the loss to be expected in AB when the scoring area of A is halved. Group SI, however, has sacrificed on three components for a relatively high $\overline{A}$. The over-all consequence of this sacrifice is reflected in $\overline{SHAB}_o$ which is less (numerically) for SI than for C, while $\overline{HAB}$ is just slightly larger for SI. To summarize, Group SI is better at A and AB, about equal in HAB and SHAB, and worse at S, H, and B. Apparently, verbal instruction has made a difference in the distribution of scoring. Whether the fundamental relationships of simultaneity or nonsimultaneity have been disturbed by the rearrangements of scoring levels must next be determined.

Analysis of independence.—This analysis can be found in the bottom two rows of Table 5. A comparison of those combinations involving A and B shows $\overline{AB}_o > \overline{AB}_p$, and very close agreement among the higher order combinations. Nonetheless, the discrepancy between AB's runs about the figure expected or a little less and does not appear disturbed by the SI Treatment. As for the other possible pairs of components, AS, BS, etc., the hypothesis of independence seemed even more reasonable for within- and between-group comparisons.

Analysis of Day 24.—On Day 24 the procedure remained unchanged except that Group SI always received B as the worst component. The consequences of this treatment can be partly assessed by comparing the results of Day 23 with Day 24 in Table 6. Component B for Group SI increases by 7 points, and A decreases by 5. No such changes show in the data of the control (C). Group SI maintains a constant value of AB across days as does Group C, but AB is greater for Group SI. The treatment for Day 24, then, illustrates two points: (*a*) knowledge of the "worst component" (the same one repeatedly) can increase a part score (B), and (*b*) this is accomplished at the sacrifice of some other component (A). Further, the consequence of events *a* and *b* was a constant level of scoring in AB from day to day.

A finer analysis of the data within Day 24 was made by examining the trial-by-trial scores. During the course of the 8 trials, Group SI showed (*a*) a progressive decline in A towards the level of A for Group C, (*b*) an increase in B to the level of B for Group C by Trial 4, and (*c*) a consistent superiority in AB over Group C. The trend of AB for Group SI was, of course, correlated with the changes in A and B. The B trend tended to augment AB, and the A trend tended to decrease the magnitude of AB. Inasmuch as AB is more

TABLE 7

EXPERIMENT 2:
OBTAINED AND PREDICTED MEANS ON DAY 24, TRIALS 1 AND 8, FOR GROUP C AND SI *

Group	Trial	Measure			
		$\overline{A}_o$	$\overline{B}_o$	$\overline{AB}_o$	$\overline{AB}_p$
C	1	51.2	81.3	44.1	41.9
	8	52.4	76.1	42.6	40.3
SI	1	68.7	74.4	53.0	51.1
	8	57.9	80.9	47.4	46.5

* $n = 36$.

sensitive to changes in A than in B (since A < B), the over-all trend of AB should be negative. The trend of AB was in line with expectation, but its complexity posed problems which defied statistical evaluation. An abbreviated set of trial data is presented in Table 7.

Finally, a test of independence for both groups on Day 24 must be made in order to determine if change in mode of responding upsets the pattern of independence. Unfortunately, no graphical recordings were available, but a test on the basis of clock scores can be made. The analysis for Trials 1 and 8 is summarized in Table 7 and shows the mean discrepancy to be about equal from group to group.

Analysis of component sums, Days 21 through 24.—Summations over certain component values of Table 6 suggest that total scoring within groups over successive days is independent of the treatments in that the sums are nearly equal from day to day. For example, consider $\sum$ ($\overline{S}$, $\overline{H}$, $\overline{A}$, $\overline{B}$) presented in Table 6. The critical comparison is, of course, between Days 23 and 24 for Group SI. A limitation of this comparison is the small amount of change in the component means.

A between-group comparison shows Group C to have the larger sum as well as the larger $\overline{SHAB}_o$. The theoretical model states that SI will have the greater $SHAB_o$ provided the sums are equal. The important point is that the *S*s of Group SI were able to redistribute their component proficiencies, but were not quite able to maintain the control sum of component proficiencies. The data of Table 6 show that the sum of the losses in S, H, and B is greater than the compensating "gain" in A.[5]

[5] Theoretically, of course, it is also possible for a gain in $\overline{SHAB}_o$ to be evidenced even with a net loss in the sum. Criterion improvement with such a loss would vary not only with the amount of decrement in Σ (S, H, A, B), but also, sum constant, with the relative sacrifices of high and low components. In the present case, the relative sacrifices are unfavorable, with B, the best performed, sacrificed less than the poorer components S and H; this is particularly true on Day 24 when all gain in A is at the expense of S and H.

The sums for $\overline{A}$ and $\overline{B}$ in Table 6 also suggest a within-group constancy from day to day, but in this comparison SI shows the larger values.

Discussion

The data have provided some leads for later and more thorough analyses. Some of the leads are discussed below.

There was no detectable, differential effect on daily mean performance of telling *S*s their actual worst components, identifying randomly selected components as worst performed, and not identifying worst components at all. This is a conclusion at variance with the general opinion that KR should be provided whenever and wherever possible in order to increase over-all proficiency. Rather, the data of Exp. 1 suggested that the KR procedures will alter the performance of particular parts and combinations of certain parts from trial to trial through a selective cueing function.

The procedures used in both Exp. 1 and 2 were effective in changing trial-to-trial scores within a day's practice. In Exp. 1, Treatments RC and LC succeeded in increasing the "lowest component" from Trial n to Trial $n + 1$. When all trials were averaged regardless of the schedule of identifying the component as LC (or RC), the mean did not appear to differ from the NC treatment. Treatments LC and RC, then, suggest more of a performance than a learning effect. The second procedure (Exp. 2) consisted of telling the *S*s to pay twice as much attention to A and to adjust the A control twice as often (unknown to all *S*s was the fact that scoring in A was made more difficult because the width of the scoring area was cut by one-half). This procedure, too, was effective

in that the experimental group was nearly able to maintain a constant level of performance in A while the control group was not. The experimental group, however, showed losses in the three other components, suggesting that constancy of proficiency across the response spectrum will be difficult to obtain from KR variations. The net loss in $\overline{AB}$ for the special instruction group, however, was substantially less than for Group C.

With B repeatedly reported as worst on Day 24, Group SI demonstrated substantial gains in $\overline{B}$, with losses in $\overline{A}$ and lesser losses in $\overline{H}$. The consequence of this multiple rearrangement of scoring profile was greater time on target in $\overline{AB}$, about equal time in $\overline{HAB}$, but a slightly lesser $\overline{SHAB}$. The changes in the part scores for the SI group represent a demonstration that variations in KR procedures can lead to relatively marked changes in performance even after much training has been administered. It might have been assumed, for example, that after 23 days of practice, habits had been established such that the pattern of responding might be relatively unsusceptible to modification.

Though combinations $\overline{HAB}$ and $\overline{SHAB}$ did not differ greatly from group to group during Days 20–24, there was some suggestion that $\overline{SHAB}$ was less for Group SI. It was previously suggested (1) that if *S*s could exchange proficiency on a well-performed component for proficiency on a less well-performed component, it was theoretically possible to raise the combination score provided the sum of indidual probabilities at remaining on target remained constant or nearly so. Group SI, however, appeared unable to maintain the required sum in four components, and thus its $\overline{SHAB}_o$ compared unfavorably with that of Group C. Though instruction and KR procedures may succeed in transferring equal amounts of attention from one component to another, it is generally unlikely that the transfer of scoring proficiency follows a simple additive equation. Specifically, for example, if the attention required to score ten units of S is applied to scoring in A, it is unlikely that A will be increased by 10 units. This would be true whenever scoring in A is more difficult than S as evaluated by performing each component alone. In the present study, it should be recalled, the scoring band of A was one-half that of S.

Thus the data of Exp. 2 have shown that in a four-part skills task, proficiency at one part may be sacrificed for additional proficiency on some other part, though the gain may not be equal to the sum of the component losses. The effects of the rearrangements are not simple, for certain first-order combinations (AB) may be raised and the third-order combinations (SHAB) may be lowered. Even so, in situations of this sort, Treatment SI could be recommended over C provided AB was a more important criterion of proficiency than SHAB.

Where the parts of a task are differentially weighted in their contribution to the magnitude of the combination score for reasons of size of scoring band (as in Exp. 2) or for other physical reasons, the effects of KR treatments may be expected to be quite complex. It is conceivable, however, that multidimensional task situations can be employed where the entire amount of proficiency sacrificed on one component may be applied to another component. Certainly further research is required to determine the conditions under which equal exchanges can or cannot take place.

The analysis of independence among components led to the same sort of finding previously reported (1), in that relatively marked departures from a chance relationship are difficult to demonstrate. The primary dimensions of response, A and B, usually led to $\overline{AB}_o > \overline{AB}_p$ by three or four percentage points, a finding reported for the Pedestal Sight Manipulation Test, and suspected to be true of the Stevens Pursuitmeter. Additional data on a Link Tester to the same effect also exist.[6] In all of these tests, the fundamental relationship between times of occurrence of scoring in two dimensions appears to be undis-

[6] Unpublished manuscript by the author.

turbed by practicing and learning over a number of days. The results from the Multidimensional Pursuit Test show that the KR procedures also did not alter the relationship between components of responding.

The present approach to the analysis of the relations among skill components and criterion combinations is suggestive of another procedure, as yet uninvestigated. The second procedure involves testing for component proficiency as *S* practices each component singly. Proficiency with two or more components practiced simultaneously might be describable in terms of proficiencies at responding to the components practiced separately.

Summary

This study was concerned with determining techniques for rearranging the rank orders and altering the magnitudes of part proficiencies within a four-part, compensatory-pursuit task. Such techniques serve two purposes: (*a*) as aids in specifying part-part and part-whole relations or composition rules, and (*b*) as potential methods for rapidly improving the criterion scores of skilled *S*s.

Two variables, instructions and knowledge of results, were used over a period of 24 days to alter *S*'s pattern of attending to the four aspects of the task. The results indicated that these variables do lead to redistributions of part proficiencies and have very predictable effects upon criterion scores. For example, identifying the worst component, say A, for *S* leads to immediate improvement in that part, and usually in a combination score such as AB. In the present data the improvement in one part was accompanied by a loss or losses in other parts such that proficiency at doing four parts simultaneously tended to fall slightly. It was suggested that these and other processes can be studied to advantage by means of the multiplication of individual component proportions of time on target.

The data failed to reveal day-by-day differences between three knowledge of results procedures which identified the worst component, a component selected at random and represented to *S* as worst, and no component at all. Since these procedures gave differential results at the trial level and no differences for aggregates of trials, it was suggested that they had momentary cue value as opposed to learning effects.

References

1. Bilodeau, E. A., & Bilodeau, I. McD. The contribution of component activities to the total psychomotor task. *J. exp. Psychol.*, 1954, **47**, 37–46.
2. Melton, A. W. (Ed.) *Apparatus tests.* AAF Aviation Psychology Program Research Report No. 4. Washington: Govt. Print. Off., 1947.
3. Melton, A. W. (Ed.) *Apparatus tests.* AAF Aviation Psychology Program Research Report No. 4. Restricted Supplement (declassified). Washington: Govt. Print. Off., 1947.

(Received November 5, 1954)